Development and Application of a Quantitative Mass Spectrometry Based Platform for Thermodynamic Analysis of Protein Interaction Networks

by

Duc Thi Minh Tran

Department of Biochemistry
Duke University

Date:_____

Approved:

_____
Michael C. Fitzgerald, Advisor

_____
Terrence G. Oas

_____
Leonard D. Spicer

_____
Edward F. Patz Jr.

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Department of
Biochemistry in the Graduate School
of Duke University

2013

ABSTRACT

Development and Application of a Quantitative Mass Spectrometry Based Platform for Thermodynamic Analysis of Protein Interaction Networks

by

Duc Thi Minh Tran

Department of Biochemistry
Duke University

Date:_____

Approved:

_____
Michael C. Fitzgerald, Advisor

_____
Terrence G. Oas

_____
Leonard D. Spicer

_____
Edward F. Patz Jr.

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biochemistry in the Graduate School of Duke University

2013

Abstract

The identification and quantification of protein-protein interactions on large scale is critical to understanding biological processes at a systems level. Current approaches for the analysis of protein-protein interactions are generally not quantitative and largely limited to certain types of interactions such as binary and strong binding interactions. They also have high false-positive and false-negative rates. Described here is the development of and application of mass spectrometry-based proteomics methods to detect and quantify the strength of protein-protein and protein-ligand interactions in the context of their interaction networks. Characterization of protein-protein and other protein-ligand interactions can directly benefit diseased state analyses and drug discovery efforts.

The methodologies and protocols developed and applied in this work are all related to the Stability of Unpurified Proteins from Rates of amide H/D Exchange (SUPREX) and Stability of Protein from Rates of Oxidation (SPROX) techniques, which have been previously established for the thermodynamic analysis of protein folding reactions and protein-ligand binding interactions. The work in this thesis is comprised of four parts. Part I involves the development of a histidine low H/D exchange protocol to facilitate SURPEX-like measurements on the proteomic scale. The histidine slow H/D

exchange protocol is developed in the context of selected model protein systems and used to investigate the thermodynamic properties of proteins in a yeast cell lysate.

In Part II an isobaric mass tagging strategy is used in combination with SPROX (i.e., a so-called iTRAQ-SPROX protocol) to characterize the altered protein interactions networks associated with lung cancer. Differential thermodynamic analyses were used to characterize the proteins in two different lung cancer cell lines; including ADLC-5M2 and ADLC-5M2-C2, in which cylcophilin A (a known protein biomarker of lung cancer) is overexpressed and knocked down, respectively. This work identified six proteins with thermodynamic stability changes in the two cell-lines.

Parts III and IV of this thesis describe the development and application of a SPROX protocol for proteome-wide thermodynamic analyses that involve the use of SILAC quantitation (i.e. **S**table **I**sotope **L**abeling by **A**mino acid in cell **C**ulture). A solution-based SILAC-SPROX protocol is described in Part III and a SILAC-SPROX protocol involving the use of cyanogen bromide and a gel-based fractionation step is described in Part IV. The SILAC-SPROX-Cyanogen bromide (SILAC-SPROX-CnBr) protocol is demonstrated to significantly improve the peptide and protein coverage in proteome-wide SPROX experiments. Both the SILAC-SPROX and SILAC-SPROX-CnBr protocols were used to characterize the ATP binding properties of yeast proteins. Ultimately, the two protocols enabled 526 yeast proteins to be assayed for binding to

AMP-PNP, an ATP mimic. A total of 140 proteins, including 37 known ATP-binding proteins, were found to have ATP binding interactions.

Dedication

This work is dedicated to my dear parents who have raised me to who I am today.

Contents

Abstract iv

List of Tables xiii

List of Figures xiv

Acknowledgements xviii

1. Introduction 1

1.1 Thermodynamic Stability, Protein Folding and Protein-Ligand Interactions 1

1.1.1 Thermodynamic Stability, Protein Misfolding and Mutagenesis 2

1.1.2 Thermodynamic Stability of Proteins reflects Protein-Ligand Binding 4

1.2 Traditional Methods for Thermodynamic Stability Measurement 6

1.2.1 Fluorescence Spectroscopy 7

1.2.2 Circular Dichroism Spectroscopy 8

1.2.3 Differential Scanning Calorimetry (DSC) 9

1.3 Protein-protein and Protein-Ligand Interactions 9

1.3.1 Analyses of Protein-Protein Interactions 10

1.3.2 Analyses of Protein - Small Molecule Interactions 10

1.3.2.1 One Small Molecule – Multiple Protein Strategy 11

1.3.2.2 One Protein – Multiple Small Molecule Strategy 15

1.4 Mass spectrometry based Proteomic Platform for Thermodynamic Analyses of Protein-Ligand Interactions 15

1.4.1 Motivation 15

1.4.2 Stability of Proteins from Rates of H/D Exchange – SUPREX 17

1.4.3 Stability of Proteins from Rates of Oxidation –SPROX 21

2. The Development of a Histidine slow HDX protocol 24

2.1 Introduction 25

2.2 Experimental Procedures 26

2.2.1 Materials 26

2.2.2 Thermodynamic Analysis of Model Protein Systems 28

2.2.3 Thermodynamic Analysis of Proteins in a Yeast Cell Lysate 30

2.2.3.1 Yeast Cell Lysate Preparation 30

2.2.3.2 Histidine HDX Analysis of Proteins from the Yeast Cell Lysate 30

2.2.4 LC-MS/MS Data Acquisition and Analysis 32

2.2.5 Histidine HDX Data Analysis 34

2.2.6 K_d Value Determination 36

2.3 Results and Discussion 37

2.3.1 Histidine HDX Protocol 37

2.3.2 Analysis of Two-State Folding Systems 41

2.3.3 Analysis of a Non-Two-State Folding System 48

2.3.4 Analysis of a Protein-Protein Interaction 50

2.3.5 Analysis of Proteins in a Yeast Cell Lysate 56

2.4 Conclusions 60

3. Application of iTRAQ-SPROX protocol to diseased state analysis in Non-Small Cell Lung Cancer 61

3.1 Introduction 61

3.1.1 Cyclophilin A and Lung cancer 61
3.1.2 The iTRAQ-SPROX protocol 62
3.2 Experimental Procedures 65
3.2.1 Cell line maintenance 65
3.2.2 Western Blot 65
3.2.3 SPROX Analysis 66
3.2.4 Methionine Enrichment 67
3.2.5 LC-MS/MS Analysis and Data Analysis 68
3.3 Results and Discussion 68
3.3.1 Western Blot data confirms Cyp-A knockdown 68
3.3.2 General Strategy 69
3.3.3 Proteomic Coverage 71
3.3.4 iTRAQ-SPROX Analysis 73
3.4 Hit Proteins Identified 76
3.5 Conclusions 78
4. Development of a SILAC-SPROX protocol and application to ATP binding discovery 80
4.1 Introduction 80
4.1.1 Motivation 80
4.1.2 Discovery of ATP-binding proteins 83
4.2 Experimental Procedures 85
4.2.1 Yeast Cell Lysate preparation 85

4.2.2 SILAC-SPROX Analysis 87
4.2.3 Proteomics Sample Preparation 88
4.2.4 LC MS/MS Analyses 89
4.2.5 Data Analysis 90
4.3 Results and Discussion 92
4.3.1 General Strategy 92
4.3.2 ATP-binding Result Summary 94
4.3.2.1 Proteome Coverage 94
4.3.2.2 Representative SILAC-SPROX data from Phosphoglycerate mutase 100
4.3.2.3 Representative SILAC-SPROX data from GAPDH 102
4.3.3 False Positive/Negative Rate 105
4.3.3.1 False Positive Rate 105
4.3.3.2 False Negative Rate 107
4.4 Conclusions 107
5. Development of a SILAC-SPROX-Cyanogen Bromide protocol and application to ATP binding discovery 109
5.1 Motivation 109
5.2 Experimental Procedures 110
5.2.1 SILAC-SPROX Analyses 110
5.2.2 Cyanogen Bromide Digestion 111
5.2.3 1-D SDS PAGE and In-gel Digestion 112
5.2.4 LC MS/MS Analyses 114

5.3 Results and Discussion 116
5.3.1 General Strategy 116
5.3.2 Proteome Coverage 120
5.3.3 Representative SILAC-SPROX-CnBr data from PGM-1 124
5.3.4 Representative SILAC-SPROX-CnBr data from Phosphoglycerate Kinase (3-PGK) 131
5.3.5 ATP binding Properties of Yeast Proteins 134
5.3.5.1 ATP binding is promiscuous 134
5.3.5.2 ATP-binding comprises of both weak and tight bindings 136
5.3.5.3 Many of ATP-binding Hits Show a Destabilization 137
5.3.6 Sensitivity of the SILAC-SPROX protocol 139
5.4 Conclusions and Future Directions 142
Appendix A 145
Appendix B 148
References 169
Biography 175

List of Tables

Table 1: Thermodynamic parameters obtained on model proteins. Values in parenthesis were previously determined by others using more conventional experimental approaches 48

Table 2: Summary of $C_{1/2}$ values determined for the histidine-containing hemoglobin peptides identified in the Hb and Hb-Hp analyses described here. "ND" indicates that no denaturant dependence was observed for the $\Delta Mass_{wt,av}$ values determined for these peptides, presumably because the histidine residues in these peptides were derived from solvent exposed regions of the intact protein structure. 51

Table 3: Summary of proteins and peptides that yielded denaturant-dependent histidine H/D exchange behavior 57

Table 4: Proteomic coverage of the iTRAQ-SPROX Cyp-A (+) vs Cyp-A (-) experiment. 72

Table 5: Normalization factors of the 8 iTRAQ reporter ions in the iTRAQ-SPROX experiment 74

Table 6: Protein hits that show changes in thermodynamic stability in the presence and absence of Cyp-A overexpression. 76

Table 7: Experiment parameters utilized in ATP-binding solution-based experiments 1A/B and 2 95

Table 8: Proteome coverage and potential protein hits from ATP-binding solution-based experiments 1A/B and 2. 96

Table 9: Experimental parameters utilized in ATP-binding gel-based experiment 1 and 2 110

Table 10: Proteome coverage from all ATP-binding experiments including solution-based 1A/B and 2; gel-based 1 and 2. 120

Table 11: Estimated K_d of proteins that have peptides showing stabilization upon binding to ATP. The K_d can only be estimated for peptides that show stabilization and if there are enough data points in the transition regions for re-construction of SPROX curves from SILAC-SPROX data. N/A means there is no calculated K_d available. 138

List of Figures

Figure 1: Schematic representation of the slow histidine H/D exchange protocol developed here 37

Figure 2: Theoretical plot showing the expected movement of $C_{1/2}$ values as a function of H/D exchange time in the slow histidine H/D exchange protocol described here. The data in the plot were generated using equation (14) and representative thermodynamic parameters of Rnase A (i.e., n =1, m-value = 3.1 kcal $mol^{-1}M^{-1}$, ΔG_f=9.2 kcal mol^{-1}, k_φ=0.288 day^{-1}). Data points at selected H/D exchange times are indicated. The dotted lines represent a linear extrapolation of data at 2.5 and 5 days exchange. The -ΔG_f Apparent" term on the y-axis represents -RT[X] where "X" is the ln-term in equation (14). 39

Figure 3: Slow histidine H/D exchange data for Rnase A. Data obtained for a His-48-containing peptide of sequence, VHESLADVQAVCSQK, is shown in (A). The solid line represents the best fit of the data to equation (13), the dotted arrow indicates $C_{1/2}$ value, the arrow labeled "1" and "2" indicates the data points for which mass spectral data is shown in (B) and (C), respectively 42

Figure 4: Slow histidine H/D exchange data obtained on a Rnase A peptide (11-25), which contained a histidine residue that was partially protected in Rnase A's three-dimensional structure. The solid line is best fit of the data to equation (13) (see text). Data points represented with open circles were excluded from the fit. 43

Figure 5: Slow histidine H/D exchange data for myoglobin using an H/D exchange time of 5 days. Data obtained on peptides containing globally protected histidine residues, His-64 and His-24 are shown in (A) and (B), respectively. Data obtained on a peptide containing partially protected histidine residues, His-81 and His-82, is shown in (C), and data obtained on a peptide containing an exposed histidine residue, His-119 is shown in (D). Peptide sequences are located at the top of each panel. The solid lines in (A) and (B) represent the best fit of each data set to equation (13) in the text. Data points represented by open circles were excluded from the fit 45

Figure 6: Slow histidine H/D exchange data obtained on myoglobin peptides, contained the globally protected histidine residues, His-64 and His-24 (respectively). The H/D exchange time was 11 days. The solid lines represent the best fit of the data to equation (13) in the text. The data point represented with an open circle was excluded from the fit 46

Figure 7: Slow histidine H/D exchange data for BCA II. Data obtained on a peptide containing histidine residues, His-118 and His-121, is shown in (A) and data obtained for a peptide containing histidine residues, His-93, His-95 and His-96 is shown in (B). The dotted arrows indicate $C_{1/2}$ values. The solid lines represent the best fit of the data to equation (13), with the data in each transition being fit separately. 49

Figure 8: Slow histidine H/D exchange data for Hb and the Hb-Hp complex. Data obtained on His-120 containing peptide from the α chain of Hb after 5 days (○ and ●) and 11 days (▽and ▼) in the presence (○ and ▽) and absence (● and ▼) of Hp is shown in (A). Similar data obtained on a His-92 containing peptide from the β chain of Hb in the presence and absence of Hp is shown in (B). The lines represent best fit of each data set to equation (13). 52

Figure 9: Representative Histidine H/D exchange data obtained from a His-120 containing peptide of superoxide dismutase 1 (SOD-1) in the presence (● and solid line) and absence (○ and dotted line) of added Zn2+. 59

Figure 10: Representative Western Blot Result of proteins from Cyp-A parental (+) and Cyp-A knockdown (-) cell line. Upper panel is an image of the ponceau stained PVDG blot preceding tranfering showing decreasing total protein amount loaded on to each lane. Lower panel is the western blot result showing decreased expression level of Cyp-A in the knockdown Cyp-A (-) cell line. 69

Figure 11: General strategy for the iTRAQ-SPROX protocol 70

Figure 12: Distribution of the iTRAQ intensities of the 113 and 121 reporter ions for un-oxidized methionine containing peptides from Cyp-A (+) on left and Cyp-A (-) samples on right. Black arrows indicate intersection of the 2 distributions (113 vs. 121). Distribution of the 114 and 119 tags are also included for comparison. 75

Figure 13: Representative iTRAQ-SPROX results from β-tubulin (A) and iEF5A (B). Bar graphs on the left represent peptides generated in Cyp-A(+) sample and bar graphs on the right represent peptides generated in Cyp-A (-) sample. Black arrow indicates estimated $C_{1/2}$ value and dotted line represents the "cut-off" line (see text) 78

Figure 14: Expected results from SILAC-SPROX solution-based experiments. (A) is an oxidized methionine peptide from a protein that has no interaction with the ligand; (B) is an oxidized methionine peptide from a protein that is stabilized by binding to the ligand; and (C) is a corresponding un-oxidized peptide of that stabilized protein. Open

circles represent data points (denaturant concentrations) that have no change in H/L ratio, closed circles represent data points that have significant H/L ratio difference. 92

Figure 15: The solution-based SILAC-SPROX protocol..... 94

Figure 16: Distribution of the $\log_2$ of the normalized H/L ratios in (A) solution-based experiment 1B and in (B) solution-based experiment 2. Dotted lines represents distribution of all peptide sequences, dash-and-dotted line represents distribution of peptides that do not contain methionine in their primary sequences and solid line represents distribution of methionine containing peptides. Inset are zoom-in image of the methionine containing peptide distributions..... 98

Figure 17: Representative SILAC-SPROX data from phosphoglycerate mutase (PGM-1) in (A) Solution-based experiment 1B, and (B) Solution-based experiment 2. Diamond shape represents data points from the un-oxidized methionine containing peptide (TVMIAAHGNSLRGLVK); square shape represents data points from the oxidized methionine containing peptide (TVM(ox)IAAHGNSLRGLVK) and triangle shape represents data points from a selected non-methionine containing peptide (LSRAIQTANIALEK)..... 100

Figure 18: SILAC-SPROX data for multiple peptides from GAPDH in Solution-based experiment 1B (A) and Solution-based experiment 2 (B). Circles are peptides with sequence NVEVVALNDPFISNDYSAYMFK; triangles are VINDAFGI-EEGLMTTVHSLTATQK; diamonds are LTGMAFRVPTVDVSVVDLTVK; and squares are (K)VVITAPSSTAPMFVMGVNEEK. Closed symbols represent un-oxidized and open symbols represent oxidized methionine containing peptides. Dotted line represents SILAC-SPROX data from a non-methionine containing peptide (VLPELQGK). 104

Figure 19: Gel-cutting strategy for (A) Gel-based experiment 1 and (B) Gel-based experiment 2. Black boxes represent relative sizes of the gel bands. Arrows indicates the estimated molecular weight ranges for each gel band. 114

Figure 20: The SILAC-SPROX-Cyanogen Bromide Protocol..... 116

Figure 21: Expected SILAC-SPROX-CnBr results from the gel-based experiments; (A) is an oxidized methionine peptide of a protein that has no interaction with the ligand; (B) is any peptides from the full-length protein that is stabilized by binding to the ligand; and (C) is any peptides from the corresponding CnBr fragments of stabilized proteins. Open circles represent data points (denaturant concentrations) that have no change in

H/L ratio, closed circles represent data points that have significant H/L ratio difference. 119

Figure 22: A comparison of the (A) proteome coverage (i.e. assayed proteins) from gel-based and solution-based experiments and (B) potential protein hits from gel-based and solution-based experiment 123

Figure 23: Distribution of known ligands for the hit proteins in this study 124

Figure 24: The sequence coverage of PGM-1 in the 2 gel-based experiments. Each arrow represents a peptide identified in the LC MS/MS. Solid arrows represent peptides identified the gel-based experiment 1. Dash arrows represent peptides identified in the gel-based experiment 2 and correspond to the full-length protein. Dotted arrows represent peptides identified in the gel-based experiment 2 and correspond to the CnBr-fragment. 125

Figure 25: Representative SILAC-SPROX-CnBr data from PGM-1 in (A) gel-based experiment 1 and (B) gel-based experiment 2. Closed circles represents data points from peptide TVMIAAHGNSLRGLVK, open circles represents data from TVM(ox)IAAHGNSLRGLVK; closed triangles represents data points from a selected non-methionine containing peptide LSRAIQTANIALEK in the CnBr Fragment; open triangles are data points from LSRAIQTANIALEK in the Full-length protein. Solid line represents peptides from gel-based exp. 1; dotted line represents peptide from gel-based exp. 2 and the CnBr fragment; dashed line represents peptides from gel-based exp. 2 and the Full-length protein (see Figure 24). 127

Figure 26: SILAC-SPROX-CnBr data of PGM-1 in gel-based experiment 2. (A) are all non-methionine containing peptides identified in Full-length protein and (B) are those identified in the 21 KDa CnBr Fragment. Dashed line and open symbols represents peptide originates from Full-length protein; dotted line and closed symbols represents peptides originates from 21 KDa CnBr Fragment. 129

Figure 27: Representative SILAC-SPROX-CnBr data from 3-PGK in gel-based experiment 2. 133

Figure 28: CnBr digestion pattern of the yeast Phosphoglycerate Kinase 1 (3-PGK). Arrows indicate CnBr cleavage sites, upper numbers indicate molecular weight of corresponding CnBr fragments. 134

Acknowledgements

Graduate school is a great place to learn and grow both scientifically and personally. I could not have done it without the help of my PhD advisor, Dr Michael C. Fitzgerald. His vision and enthusiasm in science have guided me step by step to the path of becoming an independent researcher and thinker. I owe all of my success to his belief in me and his nurturing of my scientific potential. I have learnt and grown so much in his laboratory and with that I would like to give my most sincere thanks to him.

I would like to thank my committee including Dr Terrence G. Oas, Dr Leonard D. Spicer, and Dr Edward F. Patz Jr. They have provided many valuable insights that help guided my work to the right direction.

I also would like to acknowledge my colleagues in the Fitzgerald group including Jagat Adhikari, M. Ariel Geer, Dongyu Wang, Yingrong Xu, Ryenne N. Ogburn, Julia H. Roberts, Xiaopu Jin and former group members Erin Strickland, Ying Xu, Patrick DeArmond and Graham West; who have been my terrific lab mates and friends. I would like to give special thanks to Jagat Adhikari who worked with me in part of the work presented in this dissertation. His diligence and creativity have played an important part in the success of the work.

And most importantly, I thank my parents for raising me and guide me to science and my sister who has been a role model for all of my life. I thank my friends

who have been by my side and made my time in graduate school at Duke a wonderful experience.

1. Introduction

The goal of this work is to develop a mass spectrometry based proteomics platform for the large scale analysis of protein thermodynamic stability in complex biological mixtures. Thermodynamic stability of a protein is closely linked to ligand binding and thus can be related to its function. For instance, the thermodynamic stability a protein can change upon interaction with ligand (e.g., another protein or small molecule substrate, inhibitor or co-factor). Therefore, such thermodynamic measurement of protein folding and stability are potentially useful in studying the altered protein interaction networks associated with diseases and drug actions. This introductory chapter will focus on summarizing: (i) the basic principles underlying the close relationship between thermodynamic stability and protein folding/misfolding, mutagenesis and protein-ligand interactions, (ii) the traditional methods for measurement of thermodynamic stability, (iii) the conventional methods for characterization of protein-ligand interactions, and (iv) the principles of mass spectrometry-based thermodynamic stability measurements.

1.1 Thermodynamic Stability, Protein Folding and Protein-Ligand Interactions

The study of protein folding is important because proper folding is critical for the function of the macromolecules. Understanding the folding mechanism of proteins allows researchers to replicate that process and simulate the interactions that stabilize

the native form of proteins, which directly benefits protein-drug design and discovery. The study of protein folding often includes: (i) structural characterization (ii) kinetic analyses, (iii) conformational studies, and (iv) thermodynamic analyses. Thermodynamics studies of protein folding and stability can involve measurements of the difference in free energy, in enthalpy and/or in entropy of the unfolded state with respect to folded state of the proteins, elucidating underlying physical forces that promote the folding/unfolding processes. In an effort to catalogue a large collection of all available thermodynamic stability information of proteins and mutants, Gromiha and coworkers have developed a database called "ProTherm" (Thermodynamic Database for Proteins and mutants). Protherm collects data including unfolding Gibbs free energy, enthalpy, heat capacity change, transition temperature and denaturant concentration, from literature and experiments. The database has a total of 14,500 entries and can be found at "http://www.abren.net/protherm/protherm.php". This database contains very valuable information and will facilitate the thermodynamic stability characterization of proteins.

1.1.1 Thermodynamic Stability, Protein Misfolding and Mutagenesis

Native states of proteins usually correspond to the conformations that are most thermodynamically stable under physiological conditions[1]. The thermodynamic stability of globular monomeric proteins (60 to 100 residues) often ranges from - 5 to - 15 kcal/mol. On one hand, the thermodynamic stability of proteins is sensitive to changes in

protein sequence by mutagenesis as changing in amino acid composition can alter the network of hydrophobic interactions and hydrogen bonds that stabilize the folded state of the proteins. Schellman and coworkers have demonstrated that substituting one amino acid in the primary sequence of proteins results in changes in thermodynamic stability of the mutants compared to the wild type[2]. This site-directed mutagenesis approach is useful in discovering parts of the sequence that are crucial for proper folding of proteins[2]. Recently, a study by Araya and coworkers using phage display approach has identified stabilizing mutations in a large-scale functional analysis[3]. The fundamental principle of this study is based on the concept that "Protein function is generally reduced by destabilizing mutation but can be rescued by stabilizing mutations". Using this approach Araya and coworkers have identified 15 stabilizing mutations among 47,000 mutants of the hYAP65 WW domain, one of which had more stabilizing effects than any previously known mutations[3].

On the other hand, protein misfolding can occur when a nascent chain from protein synthesis undergoes folding via partially unfolded states [4]. The abundance of these partially unfolded states and their transformation to either properly folded or misfolded proteins are determined by their relative thermodynamic and kinetic stabilities [4]. This process is tightly regulated in normal cells by various parameters including posttranslational modifications, small-molecule and protein-protein interactions. Malfunction in any steps of this folding process can lead to misfolding

which ultimately lead to a number of degenerative diseases, e.g. Alzheimer, Parkinson, Prion diseases, etc. [1, 4-5] Thermodynamic stability measurements are therefore not only important to understand the basic mechanism of folding but also to avoid misfolding related to mutagenesis and diseases.

1.1.2 Thermodynamic Stability of Proteins reflects Protein-Ligand Binding

The effect of ligand binding to proteins often includes the following (i) ligand binds at 1 site precludes occupancy of neighboring site due to space limitation and overwhelming repulsion (ii) ligand binding induces conformational changes or (iii) ligand binding drives a reaction to one side or another. Schellman and coworkers have concluded that the driving effect of ligand binding on a reaction can be quantitatively evaluated in terms of free energy of binding of the ligand to the reactants [2, 6].

In the simplest case, a protein-ligand binding interaction can be described by the following equation.

$$(1)\quad P + L \Leftrightarrow PL$$

In Equation (1), P stands for Protein, L is the Ligand and PL is the Protein-Ligand complex. Equilibrium constant K of this reaction can be defined as

$$(2)\quad K = \frac{[PL]}{[P][L]}$$

The energy of binding ΔG can be calculated by equation (3)[2]

$$(3)\quad \Delta G = -RTln(1 + K[L])$$

In equation (3), R is the gas constant (i.e. 1.986 x 10^{-3} kcal K^{-1} mol^{-1})), T is temperature in K and [L] is free ligand concentration (M). For a ligand that has multiple binding sites on the protein (n), the binding interactions can be described as below.

(4) $P + L \Leftrightarrow PL$

$P + 2L \Leftrightarrow PL_2$

…

$P + nL \Leftrightarrow PL_n$

The binding polynomial therefore can be defined using equation (5)

(5) $\sum \equiv 1 + K_1[L] + K_2[L]^2 + \cdots + K_n[L]^n$

In equation (5) K_1… K_n are the equilibrium constant of the binding reaction (4). The binding free energy for cases of multiple ligand binding sites is therefore

(6) $\Delta G = -RTln\sum$

Using this stoichiometry model of binding, Schellman suggests that it is possible to measure free energy changes induced by perturbation (i.e. mutagenesis or ligand binding) by studying the transition between the folded and unfolded state in the presence and absence of the perturbing reagents (i.e. mutations or the ligand)[2]. It is also possible to measure this binding (perturbing) energy as a function of thermal denaturation. The principle of this approach is that the equilibrium between the folded and unfolded forms of a protein (induced by thermal denaturation) is related to the change in free energy of binding, the free energy of binding in turns is a simple function of the equilibrium constant (or the K_d). The ligand typically preferentially binds to the

folded form of the proteins, hence drives the equilibrium favoring the folded states and therefore increasing the thermodynamic stability of proteins.

Recently Sanchez-Ruiz has proposed a model using the partition function approach to predict and explain the change in thermal denaturation midpoint T_m upon ligand binding[7]. This study suggests if a ligand binds to the native state of a two-state folding protein, the thermal denaturation midpoint T_m will increase as increasing ligand concentration, i.e. the protein's thermodynamic stability will increase. The ligand can only decrease thermodynamic stability of proteins when it binds to both native and unfolded states, in which case the thermal denaturation midpoint will decrease as increasing ligand concentrations. Waldron and Murphy also reported this trend of increasing T_m upon protein-ligand interactions with more complex binding models and have proposed a way to utilize thermal denaturation as an assay for drug screening[8].

1.2 Traditional Methods for Thermodynamic Stability Measurement

A typical protein refolds spontaneously with a rate constant of approximately $1s^{-1}$ and unfolds under the same condition at much lower rate (10^{-5} s^{-1}), which means at normal aqueous condition protein mainly exists in the folded states. It is harder to sample and measure the equilibrium between the folded and unfolded state when the unfolded proteins are not populated. This suggests the use of denaturant (i.e. lowering pH, increasing temperature, chemical denaturant concentration, or pressure, etc.) to facilitate the unfolding process so as to measure the equilibrium between folded and

unfolded state and ultimately to estimate the folding free energy. Any methods that can distinguish between the folded and unfolded states of proteins can be used to measure the folding equilibrium and ultimately to estimate the folding free energy of proteins. Summarized below are several spectroscopic and calorimetric approaches that are frequently used for measurement of protein thermodynamic stability.

1.2.1 Fluorescence Spectroscopy

Aromatic residues such as Tryptophan or Tyrosine have intrinsic fluorescence properties and have been exploited for monitoring the folding/unfolding equilibrium of proteins. The fluorescence properties of Tryptophan and Tyrosine are sensitive to changes in their surrounding environment occurred during the folding/unfolding process. For instance, these residues are usually buried in the hydrophobic core of proteins in which they will yield high fluorescence intensities. As the proteins unfold, these residues become exposed to solvent and their fluorescence intensities will decrease. By applying an excitation beam with wavelength of 280 nm (for both Tryptophan and Tyrosine) or 295 nm (for Tryptophan only) onto protein solutions at different conditions (i.e. different pH, temperature or chemical denaturant concentrations), one can measure the change in fluorescence intensities induced by folding/unfolding. Furthermore, the maximum emission wavelength of Tryptophan also changes upon changing in its surrounding chemical environment and can also be used to monitor protein folding/unfolding. Intrinsic fluorescence probes have been used

extensively to measure thermodynamic stability of proteins but they possess several drawbacks. First of all, not all proteins have buried Tyrosine or Tryptophan in their sequences. In order to monitor the folding/unfolding of specific domains in the proteins, Tyrosine and Tryptophan are typically introduced to the sequence by mutagenesis. As described in previous section, mutagenesis can perturb protein thermodynamic stability; therefore the observed thermodynamic stability in this case may not represent the native state of the proteins. Second of all, this approach often requires large amount of purified proteins (typically 300 μg) which can be difficult to obtain from endogenous and clinical protein samples. For these reasons the use of fluorescence spectroscopy is not suited for large-scale characterization of protein thermodynamic stabilities in complex mixtures.

1.2.2 Circular Dichroism Spectroscopy

Circular Dichroism Spectroscopy (CD) has been used extensively to study protein folding/unfolding as it has many advantages such as: (i) it provides secondary structural information, (ii) the measurement is performed in solution, and (iii) it has the capability to perform time-resolved measurement with millisecond resolution[9]. In the "far UV" CD experiment (λ ~180-260 nm), the signal represents the asymmetric backbone carbon atoms on either side of the amide bond, therefore reports on secondary structure of proteins. In the "near UV" CD experiment (λ ~ 250-350 nm), the CD signals come from disulfide bonds and aromatic residues such as Phenylalanine (250-270 nm), Tyrosine (270-290) and Tryptophan (280-300 nm). For studies of protein

folding/unfolding, the samples are typically mixed with chemical denaturant and CD signals are monitored as a function of the denaturant concentrations. The disadvantages of this technique also include the requirement of large amount of purified protein samples (typically from 1 to 10 mg/mL protein concentration) in volumes appropriate to the sample holders (i.e., cuvettes) being used.

1.2.3 Differential Scanning Calorimetry (DSC)

In the thermal denaturation of proteins, proteins are unfolded by increasing the temperature. Thermal denaturation is often coupled with Differential Scanning Calorimetry (DSC) for measurement of protein thermodynamic stabilities [10]. This method measures the change in heat capacity (C_P) of proteins from the folded (low temperature) to unfolded (high temperature) states. This method has the ability to directly measure changes in enthalpy of unfolding and melting temperature of thermal denaturation. It can also be applied to measurement of protein thermodynamic stability at different pH and in presence of different mutations [11]. This approach also has the disadvantage of requiring large amount of purified protein samples (typically 500 μL of 0.1 to 2 mg/mL).

1.3 Protein-protein and Protein-Ligand Interactions

The ability to interact and function in networks enables proteins to carry out various biological processes from enzyme regulation, biopolymer assembly, to biosynthesis pathways and signal transduction [12]. These networks of interactions

include protein-protein, protein-nucleic acid and protein-small molecule (i.e. protein-ligand) interactions. Identifying and characterizing components of these networks and quantifying their interactions will give a comprehensive understanding of biological system performance and malfunction such as in diseased conditions; and it will provide diagnostic and therapeutic targets for drug discovery efforts.

1.3.1 Analyses of Protein-Protein Interactions

Large scale analyses of protein-protein interactions is currently conducted using the yeast-two-hybrid assay, protein microarrays and tandem-affinity purification techniques coupled with mass spectrometry [12-13]. The yeast-two-hybrid assay (Y2H) [13] has good sensitivity but a major disadvantage is it can only detect binary interactions [14]. Protein micro arrays [14] and tandem-affinity purification-mass spectrometry (TAP-MS) [15] are also attractive strategies but are generally not quantitative and not able to detect binding events that result in dissociation of protein complexes. Moreover, most of these approaches usually have high false positive and false negative rates [16]. Hence, there remains an urgent need to develop alternative high throughput approaches for the identification and quantitation of protein-protein interactions in complex biological systems.

1.3.2 Analyses of Protein - Small Molecule Interactions

The characterization of protein-ligand interactions is crucial for understanding of biochemical processes and drug mode-of-action. In general the methods for Analyses

Protein-Ligand interaction can be divided into 2 categories: One small molecule-multiple proteins and one protein-multiple small molecules strategies. A recent review by McFedries and Schwaid has discussed these strategies for characterization of protein-small molecule interactions in detail [17]. The purpose of this section is to briefly mention these strategies and the advantages as well as disadvantages of the current approaches.

1.3.2.1 One Small Molecule – Multiple Protein Strategy

1.3.2.1.1 Small molecule Affinity Methods

Affinity based methods are the most common approach used to identify protein-small molecule interactions. Main principle of this approach relies on immobilization of the small molecules on a solid phase (e.g. magnetic or biotinylated beads, etc.) and affinity capturing of the proteins to the beads via interactions with the conjugated small molecules. Ong and coworkers have reported a small molecule affinity chromatography coupled with SILAC-based quantitative bottom-up proteomics approach to identify protein targets of a number of compounds[18]. In this approach, the cells are grown in "SILAC media" containing heavy Lysine and/or Arginine ($^{13}C_6$-Arginine and $^{13}C_6$,$^{15}N_2$ Lysine) resulting in 2 almost identical cell lines [18-19]. The only difference is that the proteins from SILAC labeled cell line are "heavier" than proteins from the normal cell lines; therefore they appear on different regions in the mass spectra and can be easily distinguished from their lighter counterparts. Relative abundance of the light and heavy species is determined by relative ratio between the light and heavy ion intensities. The proteins from one of the lysates (for instance the "light" one) are incubated with a

soluble form of the ligands before both lysates are subject to purification by the small-molecule coated beads. The theory is that if a protein binds to the ligand specifically, the excess amount of soluble ligand will out-compete the binding of the immobilized ligand to the proteins. Other non-interacting molecules will be washed away, non-specific binding proteins will bind indistinguishably between the two samples and the proteins with significant change in abundance in the presence and absence of excess soluble ligands will be identified as potential targets. The disadvantages of this strategy include that it requires immobilization of the ligand onto a solid support and that it is insensitive to indirect binding interactions. Conjugating the ligand to a solid surface may also perturb its native binding properties; this type of studies may not be applicable to many ligand classes and binding modes.

1.3.2.1.2 Energetics-based Proteomic Analyses

More recently, several energetics-based approaches have been developed that utilize the difference in thermodynamic stability between the ligand bound and un-bound state to detect direct and indirect protein targets of small molecules. The Drug Affinity Responsive Target Stability (DARTS) is one such approach[20]. DARTS is based on the assumption that ligand-bound and folded forms of proteins are less susceptible to proteolysis than the un-bound forms. A side by side comparison of proteolysis susceptibility of the protein mixtures in the presence and absence of the ligand of interest will reveal proteins that have differential thermodynamic stability upon binding to the ligands. The results are visualized on gel electrophoresis. Retained on the gels are

bands of folded and native proteins, protein bands that have differential intensities are identified as potential binding hits.

A similar strategy to DARTS is the Pulse Proteolysis Energetics-based approach [21]. Unlike DARTS, pulse proteolysis does not rely on proteolytic susceptibility difference of ligand-bound and un-bound proteins; rather it is analogous to more conventional chemical denaturant-induced equilibrium unfolding studies of proteins. The proteolysis step is used to distinguish between the folded and unfolded state. In pulse proteolysis, 2 aliquots of protein mixtures are equilibrated with a 3 M Urea solution, one of which was treated with ligand and the other was not. This Urea concentration is near or about the denaturant midpoint C_m of majority of proteins in the mixture. A short pulse of proteolysis by thermolysin was performed on both samples in such a way that proteolysis of folded proteins is minimal whereas most unfolded proteins are cleaved into small fragments. The remained proteins from both samples are also visualized and compared on gel electrophoresis. Folded protein bands that have differential intensities in the presence and absence of the ligand of interest will be identified as potential binding hits.

Both of the energetics-based approaches have several advantages including: (i) they are general with respect to protein and ligand classes; (ii) they are general with respect to modes of binding; (iii) they do not require derivatization of the ligands; and (iv) they can be performed on proteomics scale in a targeted manner. However they do

have several disadvantages including: (i) they relies on the ability of thermolysin to cleave the protein substrate whereas this enzyme is only active in high concentrations of Urea but not GdmCl (e.g., the enzyme is quickly inactivated in as low as 1.4 M GdmCl) [22]; (ii) in DARTS, proteins that are resistant to thermolysin cleavage cannot be assayed; (iii) in pulse proteolysis on the other hand proteins that are cleaved by thermolysin even at native condition (i.e. without Urea) are also not assayed; and (iv) both of the DARTS and pulse-proteolysis approaches rely on the resolving power and sensitivity of gel electrophoresis, which are relatively poor compared to that in LC-MS analyses. Therefore if a protein is not expressed in enough amounts so as to be visualized and successfully isolated in a spot on the gel, it cannot be assayed.

1.3.2.1.3 Chemoproteomic Target Identification

Another class of emerging techniques for measuring protein-ligand interaction is the use of chemo-reactive small molecules that can covalently attach to the active site of a class of enzymes via a bio-orthogonal chemical reaction [23]. Beside the active site reactive group, the ligand also has another reactive group that can be conjugated to an affinity chromatography apparatus using click chemistry. These special probes allow labeling of several classes of enzyme in vivo and analyses preceding cell lysis using proteomics approaches [17]. This technique is powerful however it is quite limited to the ability to synthesize such special chemical probe/substrate and not general to the protein classes and modes of binding studied.

1.3.2.2 One Protein – Multiple Small Molecule Strategy

There are multiple approaches for analyzing the binding of a target protein to multiple small molecules. One involves immobilization of the target protein onto a solid phase and then affinity capture of binding ligands from a pooled mixture of small molecules[24]. This approach relies on the assumption that the binding properties of proteins are not affected by immobilization, and requires the small molecule of interest to have special properties such as radioactivity. The other involved measuring thermal stability of proteins in the presence and absence of a pool of small molecule ligands[25]. However these approaches cannot be applied to large scale analyses of one molecule – multiple proteins.

1.4 Mass spectrometry based Proteomic Platform for Thermodynamic Analyses of Protein-Ligand Interactions

1.4.1 Motivation

As described above, the thermodynamic stability of proteins is an important property that is closely related to protein function. Changes in thermodynamic stability of proteins can indicate mutations, misfolding and/or changes in protein-protein and protein-ligand interactions. The study of protein-protein and protein-ligand interactions is not only important for characterization of diseased states but also directly benefits the protein design and drug discovery effort. Large scale measurement of thermodynamic stability of multiple proteins is a powerful technique to evaluate protein-protein and protein-ligand interactions in biological samples. Traditional approaches for

measurement of thermodynamic stability of proteins (i.e. fluorescence spectroscopy, CD spectroscopy and differential scanning calorimetry) are not suitable to large scale measurement of unpurified proteins. Current approaches for characterization of protein-small molecules interactions have made some significant progresses but are still limited to certain classes of proteins, ligands and modes of binding and are not quantitative. Energetics-based approaches (i.e. thermodynamic stability based) are powerful tools for the identification and quantification of protein-ligand interactions because: (i) they can analyze large numbers of proteins for ligand binding, (ii) they do not require the time consuming purification process of proteins, (iii) they do not require immobilization of proteins or ligands, (iv) they are general with respect to assayed protein, ligand classes and modes of binding and, (v) they are in theory quantitative.

The research in this dissertation is focused on the development of a proteomic platform for thermodynamic analysis of protein interaction networks. The methodologies developed in this dissertation represent a new energetics-based approach for the large scale analyses of protein-ligand interactions. Unlike other approaches (e.g., DARTS and Pulse Proteolysis), which are plagued by the low resolution and low sensitivity of gel electrophoresis, the new approaches described here exploit the high sensitivity and resolution of modern mass spectrometer systems that are commonly used in and quantitative bottom-up proteomic experiments.

Mass spectrometry-based proteomic analyses are widely used for the identification of proteins in biological mixtures because of the high sensitivity and high throughput capabilities [26]. The development of stable isotopic labeling strategies has also created a fast growing field of quantitative proteomics [19, 27]. However, most mass spectrometry-based proteomic studies conducted to date, have centered on analyzing changes in protein expression levels in different biological systems [27a]. The work in this thesis is focused on measuring the changes in thermodynamic stability of proteins in the context of their interaction networks. Such measurement performed on the proteins expressed in normal and diseased states is expected to add a new dimension to understanding of disease states, disease diagnosis and drug mode-of-action. This measurement will also complement the protein expression level studies that are often used to characterize disease states and drug mode-of-actions.

1.4.2 Stability of Proteins from Rates of H/D Exchange – SUPREX

SUPREX is a relatively high-throughput method for making thermodynamic stability measurements on proteins. The SUPREX method uses chemical induced folding/unfolding equilibrium of proteins in combination with amide H/D exchange and Matrix Assisted Laser Desorption/Ionization Mass spectrometry [28]. SUPREX experiments are analogous to those performed using Fluorescence or CD spectroscopy to study the chemical denaturation of proteins (see above). In SUPREX, a native and properly folded protein solution is equilibrated in a series of denaturant and D_2O

containing buffers with increasing denaturant concentrations. The chemical denaturant can be either Urea or GdmCl. The chemical denaturant induces unfolding of proteins, resulting in exposure of globally protected amide protons to solvent. Once exposed, the amide protons can exchange with deuterons in the buffers. A general scheme for this folding/unfolding and H/D exchange reaction is as followed.

$$NH_{cl} \underset{k_{cl}}{\overset{k_{op}}{\rightleftarrows}} NH_{op} \xrightarrow{k_{int}} ND_{op} \underset{k_{op}}{\overset{k_{cl}}{\rightleftarrows}} ND_{cl}$$

Scheme 1: H/D exchange reaction (adapted from reference [29]) NH_{cl} represents amide protons in the folded state of the protein (i.e. exchange incompetent); NH_{op} represents amide protons in the unfolded state of the protein (i.e. exchange competent); ND_{op} and ND_{cl} represents deuterated form of the protein in unfolded and folded state, repsectively. k_{op} represents the rate of unfolding of the protein, k_{cl} represent the rate of folding of the protein and k_{int} represents the intrinsic rate of the H/D exchange reaction.

Under EX2 exchange condition, the rate of the closing reaction k_{cl} is much faster than the rate of the H/D reaction k_{int} (i.e. $k_{cl} >> k_{int}$), therefore the observed rate of H/D exchange k_{ex} is dependent upon the folding equilibrium between the folded and unfolded state by the following equation [28].

$$(7)\ k_{ex} = \frac{k_{int}}{1+K_{fold}}$$

In the chemical induced unfolding equilibrium of proteins, the equilibrium K_{fold} is related to folding free energy and denaturant concentration by the following equation [30].

$$(8)\ K_{fold} = e^{-(\Delta G_{fold} + m[Den])/RT}$$

In equation (8), m represents the rate of change in folding free energy as changing denaturant concentration; R is the gas constant, T is temperature in K and [Den] is the denaturant concentration. Chemical induced unfolding results in increasing deuteration (i.e. increasing mass of the protein) that can be monitored by mass spectrometry as a function of denaturant concentration. The relationship between the mass gain (ΔMass) and denaturant concentration can be depicted as followed [28].

$$(9)\ \Delta Mass = \Delta M_{\infty} + (\Delta M_o - \Delta M_{\infty}) e^{-[\frac{k_{int} t}{1+K_{fold}}]}$$

In equation (9), ΔM_o is the ΔMass before the "global exchange"; ΔM_{∞} is the ΔMass after complete exchange. The dependence of the mass gain ΔMass on denaturant concentration of most two-state folding proteins is typically a sigmoidal curve with the pre-transition, transition and a post-transition region. The midpoint of this curve is indicative of thermodynamic stability of proteins. The midpoint of this curve, the $C_{1/2}$ value, is related to typical denaturation midpoint C_m by the following equation [28].

$$(10) \qquad C_{1/2} = C_m - (RT/m)\ln(\frac{k_{int} t}{0.693} - 1)$$

In equation (10), t is the exchange time. The shift in $C_{1/2}$ values is related to change in folding free energy (e.g. between ligand-bound and apo-form of proteins) by the following equation[31].

$$(11) \qquad \Delta\Delta G_{fold} = -m\Delta C_{1/2}$$

Change in folding energy can thus be used to estimate the ligand binding affinity K_d by the following equation.

$$(12) \qquad K_d = \frac{[L]}{e^{-\left(\frac{\Delta\Delta G_{fold}}{nRT}\right)}-1}$$

In equation (12), n represents the number of binding sites. As these equations suggest, the measurable ΔG_{fold}, $\Delta\Delta G_{fold}$, $C_{1/2}$ and K_d in the SUPREX experiment can be tuned by using: (i) different H/D exchange time, (ii) different temperature (or pH), or (iii) different free ligand concentrations [L]. This allows SUPREX to measure a wide variety of proteins' thermodynamic stabilities and ligand binding affinities [28].

SUPREX has many of the advantages of the energetics-based approaches described above including the ability to analyze unpurified proteins in complex biological samples. This is not only convenient because it eliminates the need to perform the protein purification step but also functionally important because native protein-protein and protein-ligand complex may be abolished following purification. SUPREX is also fast and amenable to high throughput screening of large number of ligands to a protein of interest (i.e. the one protein-multiple small molecule strategy). Recently, this protocol has been employed to screen two libraries including 1280 and 9600 compounds for binding to Cyclophilin A (i.e. a lung cancer biomarker) [32] at a rate of 6 s/ ligand [33].

Unfortunately, SUPREX is not amenable to the study of one ligand binding to multiple proteins. SUPREX requires the acquisition of protein signal in the MALDi mass spectrometry platform, which can be suppressed by the complexity of biological protein mixtures such as cell lysates. SUPREX also relies on H/D exchange reaction of amide protons that can be attenuated by back-exchange and thus is not amenable to large-scale

bottom-up proteomics, which involves prolonged protease digestion and LC MS/MS Analyses.

1.4.3 Stability of Proteins from Rates of Oxidation –SPROX

SPROX is fundamentally related to SUPREX [34]. The main difference between SPROX and SUPREX is the use of selective oxidation of methionine residues instead of H/D exchange of amide protons. SPROX inherits most of the advantages possessed by SUPREX. It also eliminates some of the disadvantages in SUPREX experiments. That is, due to the stable nature of the oxidized products, SPROX generated samples can be subject to various types of mass spectrometry analyses. For instance, SPROX generated protein samples can be analyzed using the top-down proteomics approach to measure the change in global oxidation uptake of the protein at different denaturant concentrations in various conditions. SPROX generated protein samples can also be further digested into peptides and analyzed using bottom-up proteomic approaches. The ability to interface with quantitative bottom-up proteomics is unique to SPROX and makes it a powerful tool for the analyses of protein-ligand interactions in complex biological samples.

Recently, SPROX has been coupled with isobaric mass tagging [35] for the analysis of proteins from a yeast cell lysate for binding to various ligands. The isobaric mass tags label peptides from different samples and allow combining them into a single sample

that can be analyzed in a single LC-MS/MS run. This strategy, despite having high multiplexing capability, has several drawbacks. One drawback is that it relies on the identification and quantitation of the methionine containing peptides. Unfortunately, only 20% of the identified peptides in a typical bottom-up proteomic experiment contain at least 1 methionine residue in their primary sequences (i.e. a peptide-based quantitation) [35a]. This limits the proteome coverage of the large scale analyses in SPROX experiment to 1/5 of that from typical proteomic experiments. Another drawback is the use of MS2 quantitation in the isobaric mass tagging strategy, which can be complicated by the presence of chimeric mass spectra in highly complex biological samples. The MS2 quantitation is also limited to successful selection and fragmentation of precursor ions in the data acquisition step, which can result in missing quantitation information from less abundant peptides.

Thus, there remains a need to develop new and improved mass spectrometry-based platforms to increase the scope of current SUPREX and SPROX analyses. The development of such mass spectrometry-based platforms is the main focus of this dissertation. One idea is to use another stable chemical modification that targets another residue beside methionine. This stable chemical modification should occur under relatively physiological condition for native folding of proteins (i.e. pH 7.4, Room Temperature). One such modification is the slow H/D exchange of the C_2 protons on imidazole ring of Histidine residues [36]. The development and application of this

chemical modification therefore is the focus of the first part of this work (i.e. Chapter 2 in this dissertation).

Another strategy is to improve the scope of existing SPROX methodology by increasing the number of proteins and peptides assayed. The final goal of the SPROX methodology is to perform large scale thermodynamic analysis of diseased state and drug mode-of-action in complex biological samples. To assess the validity of the current SPROX protocol for the analysis of diseased state proteome, a study is conducted and presented in Chapter 3, which involved the application of the iTRAQ-SPROX protocol to diseased state analysis in non-small cell lung cancer. The result from this study confirms the potential of SPROX in performing such large scale thermodynamic differentiation in diseased state analysis. However, several drawbacks of existing iTRAQ-SPROX protocol still remain including: (i) the high experimental errors (30-40%), (ii) the reliance on MS2 quantitation, and most importantly (iii) the pre-requisite to identify and quantify a methionine containing peptides in the bottom-up proteomics readout. Therefore the last part of this work (i.e. Chapter 4 and 5) focuses on the development of a SILAC based SPROX protocol (SILAC-SPROX) to overcome some of the drawbacks associated with current iTRAQ-SPROX protocol (i.e. the high experimental error and the reliance on MS2 quantitation in Chapter 4 and the pre-requisite to identify and quantify methionine containing peptides in Chapter 5).

2. The Development of a Histidine slow HDX protocol

Described here is a mass spectrometry-based protocol to study the thermodynamic stability of proteins and protein-ligand complexes using the chemical denaturant dependence of the slow H/D exchange reaction of the imidazole C_2 proton in histidine side-chain. The protocol is developed using several model protein systems including: ribonuclease (Rnase) A, myoglobin, bovine carbonic anhydrase (BCA) II, hemoglobin, and the hemoglobin-haptoglobin protein complex. Folding free energies consistent with those previously determined by other more conventional techniques were obtained for the two-state folding proteins, Rnase A and myoglobin. The protocol successfully detected a previously observed partially unfolded intermediate stabilized in the BCA II folding/unfolding reaction; and it could be used to generate a K_d value of 0.24 nM for the Hb-Hp complex. The compatibility of the protocol with conventional mass spectrometry-based proteomic sample preparation and analysis methods was also demonstrated in an experiment in which the protocol was used to detect the binding of Zn^{2+} to superoxide dismutase in the yeast cell lysate sample. The yeast cell sample analyses also helped define the scope of the technique, which requires the presence of globally protected histidine residues in a protein's three-dimensional structure for successful application.

2.1 Introduction

The utility of slow histidine H/D exchange as probe of protein structure has been demonstrated in continuous labeling experiments [11, 17, 36]. In the histidine H/D exchange protocol described here, the denaturant dependence of the H/D exchange reaction is probed in order to evaluate the more global thermodynamic parameters associated with the more global unfolding/refolding reactions in proteins and protein-ligand complexes. The protocol developed here is similar to that used in the SUPREX technique [28] (see Introduction), which exploits the amide H/D exchange reaction in proteins.

The half-life of the H/D exchange reaction of an unprotected histidine residue is on the order of ~2 days [36], which is considerably longer (~400,000 times longer) than that of the H/D exchange reaction of an unprotected amide proton [6b]. This means that the minimum H/D exchange time required in the histidine H/D exchange protocol is much longer than that in SUPREX, as both protocols require the use of H/D exchange times that are at least 2.5 times the half-life of the H/D exchange reaction of an unprotected site. Thus, H/D exchange times of at least 5 days are required in the histidine H/D exchange protocol, whereas H/D exchange times on the order of minutes to hours can be employed in SUPREX. It also means that the extent of back-exchange during the mass spectral sample preparation and analysis is relatively small, even when conventional proteomic sample preparation and analysis methods are used. Thus,

unlike SUPREX, the histidine H/D exchange protocol developed here can be interfaced with standard mass spectrometry-based proteomics platforms.

As part of this work the histidine H/D exchange protocol is developed and applied to a series of model protein systems including: ribonuclease (Rnase) A, myoglobin, bovine carbonic anhydrase (BCA) II, hemoglobin (Hb), and the hemoglobin-haptoglobin (Hb-Hp) complex. The compatibility of the protocol with conventional mass spectrometry-based proteomics sample preparation and analysis methods is evaluated in an experiment in which the protocol is applied to the proteins in a yeast cell lysate sample both in the absence and in the presence of added Zn^{2+} in order to test the ability of the protocol to detect the binding of Zn^{2+} to unpurified superoxide dismutase. The results obtained on proteins in the yeast cell lysate samples also help define the scope of the technique, which relies on the presence of at least one globally protected histidine residue in a protein's three-dimensional structure for successful analyses.

2.2 Experimental Procedures

2.2.1 Materials

The following materials were purchased from Sigma-Aldrich (St. Louis, MO): Rnase A from bovine pancreas (≥ 60 wt. %), myoglobin from equine skeletal muscle (≥ 95 wt. %), BCAII from bovine erythrocytes (≥ 80 wt. %), trypsin from porcine pancreas (proteomic grade), deuterium oxide (D_2O; 99.9 atom % D), sodium deuteroxide (35 wt. % in D_2O, 99.9 atom % D), deuterium chloride (20 wt. % in D_2O, 99.9 atom % D), 2-(N-

morpholino)ethanesulfonic acid (MES), Tris(2-carboxyethyl)phosphine hydrochloride (TCEP●HCl), and *S*-methyl methanethiosulfonate (MMTS). Guanidine hydrochloride (GdmCl) and acetone were purchased from EMD Chemicals, Inc. (Gibbstown, NJ). Deuterated GdmCl was prepared by dissolving the GdmCl purchased from EMD Chemicals, Inc. in D_2O and lyophilizing the sample to dryness. Four dissolution/lyophilization cycles were used to maximize the deuterium content of the GdmCl. Immobilized porcine trypsin and immobilized bovine chymotrypsin (sequencing grade) were purchased from Princeton Separation (Adelphia, NJ). Dithiothreitol (DTT), iodoacetamide (IDA) and acetonitrile (ACN) were purchased from Fisher Scientific (Pittsburgh, PA). Formic acid (FA) was purchased from Thermo Scientific (Rockford, IL) and the Amicon Ultra-0.5 mL Centrifugal Filter units with a 10,000 molecular weight cut-off were purchased from Millipore (Billerica, MA).

Stroma free hemoglobin (SFH) was obtained by precipitation by ammonium sulfate followed by a step where the endogenous phosphates are removed and then the protein is purified by FPLC following a previously published method.[37] Hp, predominantly of the 1-1 phenotype, was purchased from Sigma-Chemical (St. Louis, MO) and was 99% pure. For complex formation, Hb was mixed with Hp in a 1:1 ratio (protein concentration). The complex was further purified using a BioSep 3000 column from Phenomenex (Torrence, CA) using a 50 mM potassium phosphate buffer (pH 7.4).

Spectroelectrochemistry and MALDI-MS were used to verify the Hb-Hp complex formation as described previously [37].

2.2.2 Thermodynamic Analysis of Model Protein Systems

Histidine H/D exchange analyses on Rnase A were initiated on a 1 mM stock solution of fully protonated protein prepared in H_2O. A 10 mL aliquot of the stock solution was diluted into 90 mL of deuterated 20 mM sodium phosphate buffer (pD 7.4) containing 20 mM sodium acetate, 100 mM NaCl, and the following concentrations of deuterated GdmCl: 0.0, 0.6, 1.2, 1.7, 2.2, 2.5, 3.0, 3.2, 4.3, and 5.3 M. The samples were incubated for 5 days at 37 ^{o}C. The H/D exchange reactions in each buffer were quenched by addition of 10 μL of FA before addition of 1 mL of acetone to precipitate protein in each denaturant-containing buffer. The protein precipitation step was allowed to proceed overnight at -20 ^{o}C before the samples were centrifuged at 20,000 x g for 30 min to pellet the insoluble protein. The resulting protein pellets were dissolved in 50 μL of 0.1 M ammonium bicarbonate buffer (pH 8) containing 0.1% SDS. The samples were reduced with 10 mM DTT at 80 ^{o}C for 15 min, and alkylated with 20 mM IDA in the dark for 30 min. The protein was digested with a combination of 0.5 μg immobilized trypsin and 0.5 μg immobilized chymotrypsin for 1 hr. each at 37 ^{o}C and gentle shaking. The samples were centrifuged and the immobilized proteases were separated from the supernatant, which was adjusted to pH 2-3 with 10% TFA. The SDS in each supernatant was removed using SDS removal tips (PolyLC, Columbia, MD) according to the

manufacturer's protocol. The resulting samples were subjected to LC-MS/MS analyses as described in section 2.2.4 below.

Histidine H/D exchange analyses on myoglobin, BCAII, Hb, and the Hb-Hp complex were initiated on 0.5 to 1 mM stock solutions of these proteins that were prepared in H_2O. In each case a 10 μL aliquot of stock solution was diluted into 90 μL of deuterated 50 mM MES buffer (pD 7.4) containing 50 mM NaCl and the following GdmCl concentrations: 0, 0.7, 1.2, 1.8, 2.2, 2.4, 2.7, 2.9, 3.2, 3.6 and 4.2 M (for myoglobin and BCAII), and 0, 1.4, 1.8, 1.9, 2.2, 2.4, 2.9, 3.4, 4.0, 4.5, 5.3 M (for Hb and Hb-Hp). The samples were incubated at 37 ^{o}C for either 5 or 11 days. After the specified H/D exchange time, a 50 μL aliquot of each H/D exchange reaction was removed and diluted into 350 μL H_2O. Each sample was added to an Amicon Ultra-0.5 mL Centrifugal Filter unit and centrifuged for 15 min at 14,000 x g. Two 200 μL aliquots of 1 mM Tris buffer (pH 8.5) containing 8 M Urea (referred to as UA solution) were sequentially added to each filter unit and then centrifuged for 15 min each at 14,000 x g. Aliquots of 100 μL of UA containing 20 mM TCEP●HCl or 40 mM MMTS were subsequently added to reduce and alkylate the protein samples in the filter units for 1 hr. at RT. Three 100 μL aliquots of UA solution were sequentially added to the filter units and centrifuged for 15 min each at 14,000 x g. The protein in each filter unit was reconstituted in 120 μL of 50 mM ammonium bicarbonate buffer (pH 8.0) before the sample were digested in the filter unit with ~ 2 μg of trypsin for 12-15 hrs.; and with 1 μg of chymotrypsin for 1 hr. at room

temperature. The resulting samples were each subjected to a single LC-MS/MS analysis as described in section 2.24 (see below).

2.2.3 Thermodynamic Analysis of Proteins in a Yeast Cell Lysate

2.2.3.1 Yeast Cell Lysate Preparation

The yeast overexpression strain was grown overnight in 50 mL of a 2% raffinose SC-Ura solution. The cell culture was transferred into 500 mL of 2% raffinose SC-Ura to give an $O.D._{600}$ of 0.3 and incubated at 30 ^{o}C until log phase ($O.D._{600}$ of 0.8 ~ 1.2). The overexpression of cyclophilin A was induced upon the addition of 250 mL of 6% galactose in 3xYP solution. After six hours of induction 125 mL fractions of the culture were pelleted by centrifuging at 1000 x g for 10 min at 4^{o}C. The cells in a pellet were lysed in 20 mM phosphate buffer (pH 7.4) containing 1x concentration of Halt protease inhibitor cocktail (Thermoscientific, Rockford, IL). Cell lysis was accomplished by vigorously shaking the sample with glass beads (0.5 mm diameter) at RT a total of ten times for 20 sec each with a 1-min-incubation on ice between shakings. The resulting suspension was centrifuged at 14,000 x g for 10 min. The supernatant was used in the two yeast cell lysate experiments described below.

2.2.3.2 Histidine HDX Analysis of Proteins from the Yeast Cell Lysate

The yeast cell lysates, which typically contained ~10 mg/ml of total protein, were prepared as previously described using a cyclophilin A (YDR155C) overexpression strain that was purchased from Open Biosystems, Huntsville, Al, USA.[38] More detailed

information about the yeast cell lysate preparation is provided in the Supporting Information. In the first yeast cell lysate experiment, 10 μL of the yeast cell lysate, which contained about 100 μg of total protein, was diluted into 90 μL of deuterated 20 mM phosphate buffers (pD 7.4), containing the following GdmCl concentrations: 0.5, 0.8, 1.1, 1.6, 2.0, 2.9, 2.5, 3.6, 4.0, or 4.5 M. The protein samples in each buffer were subject to the same histidine H/D exchange analysis as described above for Rnase A except that protease digestion step was allowed to proceed for 6 hr. and it only involved one protease, trypsin, with ratio of 1:20 enzyme to substrate ratio.

In the second yeast cell lysate experiment, the yeast cell lysate was divided into two 135 μl fractions, with each fraction containing ~ 1350 μg of total protein. A 15 mL aliquot of a 1 mM $ZnCl_2$ solution was added to one fraction and 15 ml of H_2O was added to the second fraction. Both fractions were incubated at room temperature for 15 min, before 10 mL aliquots of each fraction were diluted into 90 μL of deuterated MES buffers (pD 7.4) containing 50 mM NaCl and the following GdmCl concentrations: 0.0, 1.2, 1.4, 1.9, 2.2, 2.4, 2.7, 2.9, 3.2, 4.2, and 5.3 M. The protein samples in each buffer were subject to the same histidine H/D exchange analysis as described above for Rnase A except that after the acetone precipitation, the protein pellets were dissolved in freshly prepared 50 mM ammonium bicarbonate buffer (pH 8.0) containing 6 M urea before they were reduced with 10 mM DTT at 80 ^{o}C for 15 min, alkylated with 20 mM IDA for 30 min in

the dark, diluted 5-fold with 50 mM ammonium bicarbonate buffer (pH 8.0), and digested with 1:20 trypsin for ~ 6 hrs.

2.2.4 LC-MS/MS Data Acquisition and Analysis

LC-MS/MS analyses were performed on an Agilent 6520 Q-TOF mass spectrometer equipped with a Chip Cube interface. Protease digests of the purified protein samples in this work were analyzed on an HPLC Chip containing a 43 mm x 75 μm reversed-phase HPLC column with 5 μm Zorbax 300SB-C18 packing material. The elution gradient used for analysis of the protease digests of the purified proteins samples was 3 to 70% Buffer B over 7 min. Buffer A was 0.1% FA in H_2O and buffer B was 0.1% FA in ACN. The flow rate was 0.4 μL/min. The capillary voltage was set to 1800 V. The flow rate of the drying gas (i.e. N_2) was maintained at a temperature of 350 ^oC, was 6 L/min. The skimmer and fragmentor voltages were set at 65 and 175 V, respectively. The collision energy used in the MS/MS experiments was either 3.5V/100 Da with a -4.8V offset (i.e., for the Rnase A experiments and for the first yeast cell lysate experiments), 3.6V/100 Da with a -4.8V offset (i.e., for the myoglobin, BCAII, hemoglobin, and hemoglobin-haptoglobin experiments), or 3.9V/100 Da with a 2.9 V offset (i.e., for the second yeast cell lysate experiments). The inclusion window width for precursor ion selection in the MS/MS experiments was 4 m/z. The scan rate in the MS/MS experiments was 4 precursor spectra/sec and 3 product spectra/sec. A total of 3 precursor ions were selected for fragmentation per cycle.

LC-MS/MS analyses of the protease digestion products generated from the yeast cell lysate samples were identical to that described above for the protease digests of the purified protein samples with a few exceptions. One exception was that the HPLC Chip contained a 150 mm x 75 um HPLC column and the gradient used for the analysis was 2% B over 5 min, 2 to 15% B over 2.5 min, 15 to 45% B over 78 min, and then 45 to 100% B over 10 min at 0.4 μL/min. The other exceptions were that the capillary voltage was set to 1875 V, the scan rate in the LC-MS/MS experiments was 3 precursor spectra/sec and 2 product spectra/sec, and 4 precursors per cycle were selected for fragmentation.

LC-MS/MS data were analyzed using Spectrum Mill Proteomics Workbench software (Rev. A03.03.084) from Agilent Technologies, Inc. (Santa Clara, CA). In analyses using the Spectrum Mill software, the precursor and product ion mass tolerance were either 20 ppm (i.e., for analysis of the Rnase A and yeast cell lysate data), 50 ppm (i.e., for the BCA II and myoglobin data analyses) or 15 ppm (i.e., for the hemoglobin and hemoglobin-haptoglobin data analyses). The protease was set to trypsin (i.e., for the yeast cell lysate data analysis), Trypsin/Chymotrypsin (i.e., for the hemoglobin and hemoglobin-haptoglobin data analyses), or no enzyme (i.e., for the Rnase A, BCAII, and myoglobin analyses). The maximum ambiguous precursor charge was set to 7. The carbamidomethylation (i.e., in the Rnase A and yeast cell lysate data analyses) or MMTS (i.e., in the myoglobin, BCAII, Hb and Hb-Hp data analyses) modification of cysteine was set as a fixed modification, and oxidation of methionine

was set as variable modification. The databases used for the purified protein analyses were comprised of the known amino acid sequences of the protein being analyzed (e.g., Rnase A, myoglobin, BCAII, or hemoglobin). The database used for the lysate experiment was the *Saccharomyces cerevisiae* subset from NCBinr (downloaded from ftp://ftp.ncbi.nlm.gov/blast/db/FASTA on June 20th 2009).

2.2.5 Histidine HDX Data Analysis

The deuterium content of histidine-containing peptides was determined by calculating the weighted average mass of the five most abundant isotopologues observed for each peptide in the mass spectra obtained in the LC-MS/MS analyses. This weighted average mass value was used to calculate a $\Delta Mass_{wt,av}$ value of a given histidine-containing peptide at each denaturant concentration by subtracting the theoretical weighted averaged mass value expected for the fully protonated peptide, which was also calculated using the five most abundant isotopologues expected for each peptide. The $\Delta Mass_{wt,av}$ values obtained in this work were determined using the summed ion intensities from the approximately 5 mass spectra that were collected on a given peptide in each LC-MS/MS analysis. The $\Delta Mass_{wt,av}$ values were plotted as a function of the denaturant concentration. The data in the $\Delta Mass_{wt,av}$ versus [GdmCl] plots were fit to a standard sigmoidal equation, equation (13), using a non-linear least-squares analysis based on the Marquardt-Levenberg algorithm in Sigma Plot in order to

evaluate the concentration of denaturant at the transition midpoint of the resulting sigmoidal curve (i.e., the $C_{1/2}$ value).

$$(13) \qquad y = y_o + \frac{a}{1+e^{-\left(\frac{x-C_{1/2}}{b}\right)}}$$

In equation (13), x was the [GdmCl], y was the ΔMass value, y_0 was the ΔMass value in the pre-transition region, a was the amplitude of the transition, and b was related to the steepness of the transition. Ultimately, the $C_{1/2}$ value was used to calculate a folding free energy according to equation (14).

$$(14) \qquad \Delta G_f = -mC_{1/2} - RT\left[\ln\left(\frac{\frac{k_\varphi t}{0.693}-1}{\frac{n^n}{2^n}[P]^{n-1}}\right)\right]$$

In equation (14), which was previously been reported for the analysis of SUPREX data [31], ΔG_f is the folding free energy of the protein, k_φ is first order rate constant of the slow H/D exchange reaction at the C_2 position on an unprotected histidine imidazole side chain, m is $\delta\Delta G_f/\delta[GdmCl]$, T is the temperature, R is the ideal gas constant, t is the H/D exchange time, and [P] is the protein concentration expressed in n-mer equivalents. In all the ΔG_f value calculations described here using equation (14), T was 310 K, k_φ was set at a value of 0.288 day^{-1} (based on the data in reference [36]), and n was 1 with exception of the Hb analyses in which n was 2 (as the Hb tetramer has been shown to dissociate into to a/b dimers in other GdmCl-induced equilibrium unfolding experiments [39]). In the Rnase A and myoglobin analyses, previously determined m-values of 3.1 [40] and 3.71 [41] kcal mol^{-1} M^{-1} (respectively) where used in equation (14) for

the ΔG_f calculations. In the transition midpoint analysis method used to analyze the Hb and Hb-Hp data, the $C_{1/2}$ values obtained at the different H/D exchange times were fit to equation (14) using a linear least squares analysis in which the y-intercept and slope of the best-fit line were taken as the ΔG_f value and *m*-value, respectively.

2.2.6 K_d Value Determination

The K_d value of the Hb-Hp complex was calculated using equation (15):

$$(15) \qquad K_d = \frac{4L_{total}e^{-\frac{\Delta\Delta G_f}{NRT}} - 4P_{total}(e^{-\frac{\Delta\Delta G_f}{NRT}} - 1)}{\left(2e^{-\frac{\Delta\Delta G_f}{NRT}} - 1\right)^2 - 1}$$

In equation (15), the derivation of which has been previously described,[34] L_{total} is the concentration of ligand and P_{total} is the concentration of protein, N is the number of independent binding sites, R is the gas constant, T is the temperature in Kelvin, and $\Delta\Delta G_f$ is the binding free energy. The binding free energy was calculated from the ΔG_f values obtained for Hb in the absence and in the presence of Hp.

2.3 Results and Discussion

2.3.1 Histidine HDX Protocol

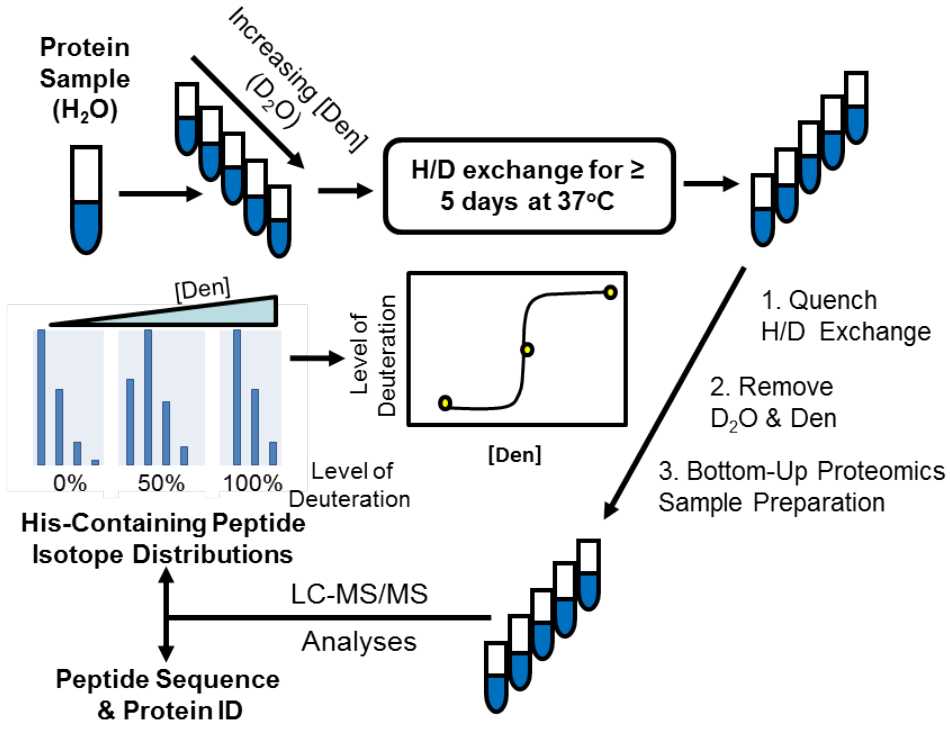

Figure 1: Schematic representation of the slow histidine H/D exchange protocol developed here

The protocol developed here (Figure 1) involves dilution of a protein sample into a series of deuterated buffers containing increasing concentrations of a chemical denaturant (e.g. GdmCl). The protein samples in each deuterated buffer are allowed to undergo H/D exchange at 37 °C and pD 7.4 for the same amount of time. The H/D exchange time (t in equation (14)) should be at least 5 days, which is equivalent to ~2.5 half-lives of the H/D exchange reaction of a C_2 proton in the imidazole side-chain of an

unprotected histidine residue [36]. This is necessary to ensure that the ln-term in equation (14) is >>0. It is also important that the H/D exchange time not be so long as to compromise the integrity of the protein sample (e.g., the protein sample can be oxidized and/or degraded with proteases if they are present). However, we note that no such problems were observed when H/D exchange time between 5 and 11 days were used to analyze the model proteins in this work.

An important consideration in the development of the histidine H/D exchange protocol described here was the choice of H/D exchange time. An H/D exchange time of at least 5 days is required for the protocol to produce reasonably accurate ΔG_f values using equation (14). This is because the $k_\varphi t$ term in equation (14) must be significantly greater than 0.693 or the accuracy of the linear extrapolation method employed in our data analysis is compromised (see Figure 2). For example, the use of a 2.5 day H/D exchange time would lead to the calculation of an aberrantly low ΔG_f value (see Figure 2). In theory, data collected using longer H/D exchange times should produce the most accurate ΔG_f values (see Figure 2).

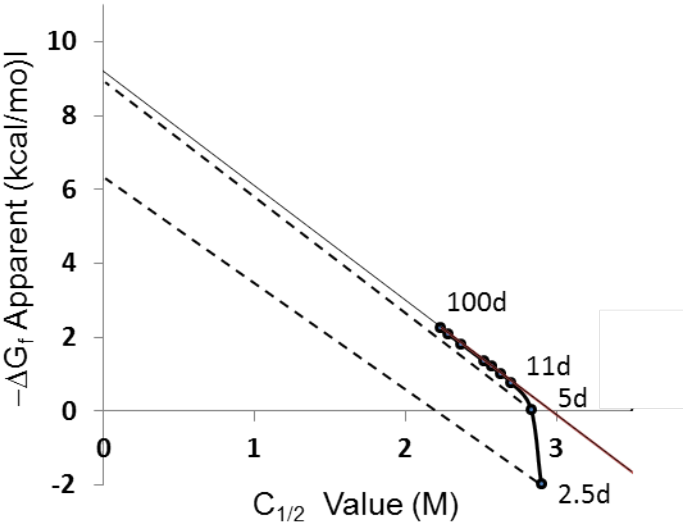

Figure 2: Theoretical plot showing the expected movement of $C_{1/2}$ values as a function of H/D exchange time in the slow histidine H/D exchange protocol described here. The data in the plot were generated using equation (14) and representative thermodynamic parameters of Rnase A (i.e., n =1, m-value = 3.1 kcal mol^{-1}M^{-1}, ΔG_f =9.2 kcal mol^{-1}, k_φ=0.288 day^{-1}). Data points at selected H/D exchange times are indicated. The dotted lines represent a linear extrapolation of data at 2.5 and 5 days exchange. The -ΔG_f Apparent" term on the y-axis represents -RT[X] where "X" is the ln-term in equation (14).

Unfortunately, the use of long exchange times (e.g., greater than several weeks) can not only be impractical but also can potentially compromise the integrity of the protein sample (e.g., the protein sample could be degraded and/or oxidized). Protein degradation may be especially problematic in unpurified protein samples (e.g., cell lysates) where proteases may be present. It is possible that such problems may even manifest themselves when 5 and 11 day H/D exchange times are used to analyze some samples. However, such complications with sample degradation are likely to be mitigated in protein ligand binding experiments where differential measurements are

made on the same sample using the same H/D exchange time. Moreover, potential problems can be identified if experiments are done using both 5 and 11 day H/D exchange times. If significantly different thermodynamic parameters are determined using the data at the 5 and 11 day H/D exchange time, this would signal a potential problem. We note that no such problems were observed using the model systems in this work.

In the H/D exchange reactions all the labile hydrogens in a protein are subject to exchange including the relatively slow exchanging C_2 protons in the imidazole side chain of histidine residues, the fast exchanging amide hydrogens, and the even faster exchanging side-chain hydrogen atoms bonded to nitrogen, oxygen and sulfur. Ultimately, the H/D exchange reactions are quenched by acidifying the solution and lowering temperature. The protein samples in the denaturant containing-buffers are each subjected to a desalting step in which the denaturant and the D_2O are removed using spin columns, acetone precipitation, or TCA precipitation.

After the desalting step, the protein samples are reduced, alkylated, and digested with a proteolytic enzyme (e.g., trypsin) according to standard mass spectrometry-based proteomic protocols. During these sample handling steps the protein samples are denatured and subject to alkaline pH conditions for extended periods of time (e.g., 6 to 12 hrs.). The H/D exchange rates of unprotected amide and side-chain protons/deuterons under these conditions are >1 s^{-1} [6b]. Therefore, deuterons that were

incorporated into the peptide backbone and the amino acid side chains during the initial H/D exchange reaction are nearly all exchanged back to protons. However, the large majority of deuterons that were exchanged into the C_2 position of the imidazole side chain of histidine residues are not back-exchanged to protons because of the longer half-life required for back-exchange. The resulting peptides are subjected to a mass spectrometric analysis in which the peptides are sequenced to identify histidine-containing peptides and the mass spectral data are used to determine a $\Delta Mass_{wt,av}$ value for each histidine-containing peptide at each denaturant concentration in which the H/D exchange reaction was performed on the intact protein. Ultimately, the $\Delta Mass_{wt,av}$ values obtained for a given histidine-containing peptide are plotted as a function of the denaturant concentrations, and the data are used to calculate a protein folding free energy as described previously.

2.3.2 Analysis of Two-State Folding Systems

Rnase A, which contains four histidine residues (His-12, His-48, His-105, and His-119), was initially analyzed using the above protocol. Based on the results of earlier histidine H/D exchange studies of Rnase A in which the time course of histidine H/D exchange was studied [36], His-105 and His-119 in Rnase A are solvent exposed, His-12 is partially protected, and His-48 is buried in the hydrophobic core of Rnase A's native three-dimensional structure. The slow histidine H/D exchange data obtained here for one of the detected His-48 containing peptides is shown in Figure 3.

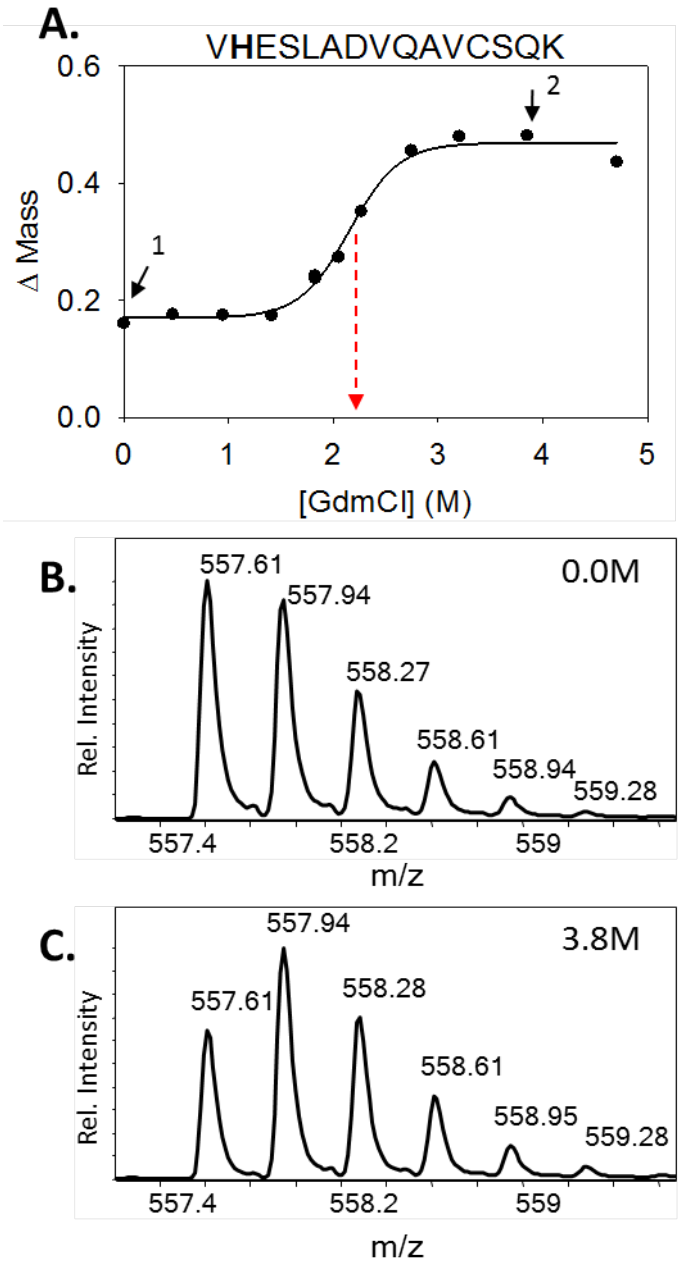

Figure 3: Slow histidine H/D exchange data for Rnase A. Data obtained for a His-48-containing peptide of sequence, VHESLADVQAVCSQK, is shown in (A). The solid line represents the best fit of the data to equation (13), the dotted arrow indicates $C_{1/2}$ value, the arrow labeled "1" and "2" indicates the data points for which mass spectral data is shown in (B) and (C), respectively

As expected for a globally protected histidine residue, there was a clear denaturant dependence to the H/D exchange behavior of His-48 (Figure 3A). A His-12 containing peptide showed a similar denaturant dependence to its H/D exchange behavior but the curve had smaller amplitude (see Figure 4 below).

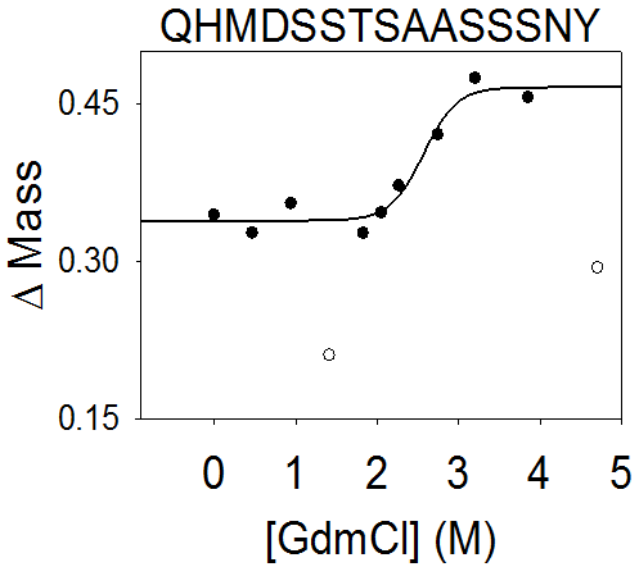

Figure 4: Slow histidine H/D exchange data obtained on a Rnase A peptide (11-25), which contained a histidine residue that was partially protected in Rnase A's three-dimensional structure. The solid line is best fit of the data to equation (13) (see text). Data points represented with open circles were excluded from the fit.

Visual inspection of the isotopologues obtained here for two other Rnase A peptides, Rnase A (105-115) and Rnase A (106-120) covering His-105 and His-119, indicated that these histidine residues were each ~50% deuterated and that the extent of deuteration was unchanged with denaturant concentration. Such H/D exchange behavior is expected for peptides containing these histidine residues, which were from

solvent exposed regions of protein structure (see discussion of expected deuteration levels below).

Expected Deuteration Levels. In theory, the $\Delta Mass_{wt,av}$ value for a completely exchanged histidine-containing peptide should be close to 0.7 Da, considering the deuterated buffers were 90% D_2O and the H/D exchange reaction was allowed to proceed for 5 days (i.e., 2.5 half-lives of the H/D exchange reaction of an unprotected histidine residue) (i.e., 90% of 76% complete). However, the post-transition baseline $\Delta Mass_{wt,av}$ values observed here were typically ~ 0.5 per histidine residue. Presumably, a fraction (~30%) of the deuterons that exchanged into the C_2 position of the histidine residues were back-exchanged with protons in protonated buffer during the mass spectrometry sample preparation and analysis steps of the protocol. However, it is important to note that as long as the extent of this back exchange reaction is similar for each data point that defines the curve, the $C_{1/2}$ value determination is not compromised.

In the histidine H/D exchange analysis of myoglobin, peptides covering 9 of the 11 histidine residues in myoglobin's amino acid sequence were identified in the mass spectral readout. These 9 histidine residues included 2 that are globally protected (His-24 and His-64), 5 that are partially protected (His-36, His-81, His-82, His-113, His-116), and 2 that are exposed (His-48 and His-119), based on the slow histidine H/D exchange behavior of the peptides in this work (see below) and on a visual inspection of myoglobin's three-dimensional structure [19].

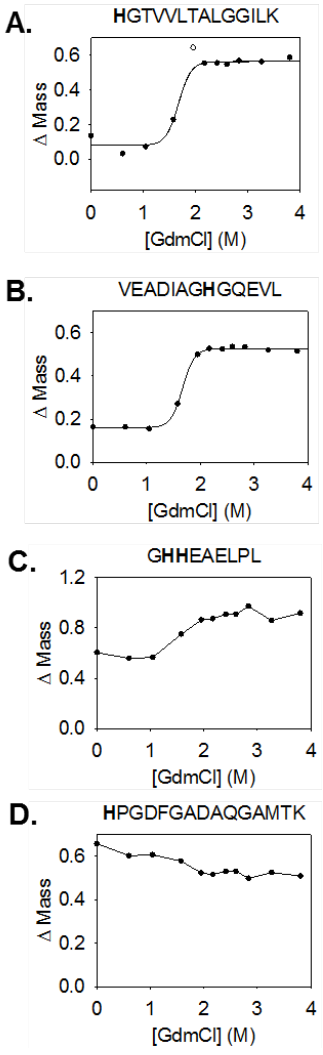

Figure 5: Slow histidine H/D exchange data for myoglobin using an H/D exchange time of 5 days. Data obtained on peptides containing globally protected histidine residues, His-64 and His-24 are shown in (A) and (B), respectively. Data obtained on a peptide containing partially protected histidine residues, His-81 and His-82, is shown in (C), and data obtained on a peptide containing an exposed histidine residue, His-119 is shown in (D). Peptide sequences are located at the top of each panel. The solid lines in (A) and (B) represent the best fit of each data set to equation (13) in the text. Data points represented by open circles were excluded from the fit.

The data obtained on peptides containing the two protected histidine residues, His-24 and His-64, showed a clear denaturant dependence to their H/D exchange behavior and yielded $C_{1/2}$ values that were 1.7 M and 1.6 M [GdmCl], respectively (see Figure 5A and 5B). The data obtained on peptides containing a partially protected histidine residue could also be fit to a sigmoidal curve, however, the amplitude of the curve was small (see Figure 5C). Peptides containing an exposed histidine residue (e.g., His-119) were essentially all exchanged at each denaturant concentration (see e.g., Figure 5D).

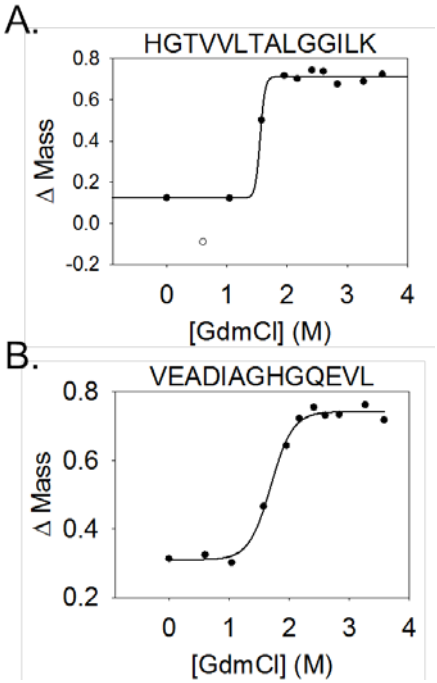

Figure 6: Slow histidine H/D exchange data obtained on myoglobin peptides, contained the globally protected histidine residues, His-64 and His-24 (respectively). The H/D exchange time was 11 days. The solid lines represent the best fit of the data to equation (13) in the text. The data point represented with an open circle was excluded from the fit.

The histidine H/D exchange protocol described here utilizes a peptide-readout. However, the $C_{1/2}$ values generated for the histidine-containing peptides detected in the mass spectral readout are representative of the protein folding unit from which they derive. In the case of a two-state folding globular protein such as Rnase A and myoglobin, each protein molecule is considered to be a single folding unit,[41] and therefore the $C_{1/2}$ values derived from different globally-protected histidine-containing peptides in these proteins are expected to be similar and to yield folding free energies that are comparable to those determined for the intact protein using other techniques. Indeed, the $C_{1/2}$ value measured here for the Rnase A(47-62) peptide can be used in equation (14) to calculate a ΔG_f value of -6.9 kcalmol^{-1} for Rnase A, which is within 30% of that previously determined[40] for this protein (see Table 1). The similar $C_{1/2}$ values obtained from the two different globally protected histidine-containing peptides in the myoglobin analysis (see Table 1) are consistent with a two state folding mechanism. The ΔG_f values calculated using equation (14) and these $C_{1/2}$ values are also similar and within 20% of that previously determined [41] for myoglobin (see Table 1).

Table 1: Thermodynamic parameters obtained on model proteins. Values in parenthesis were previously determined by others using more conventional experimental approaches

Protein Peptide	$C_{1/2}$ (M)	m (kcal $mol^{-1}M^{-1}$)	ΔG_f (kcal mol^{-1})
Rnase A			
VHESLADVQAVCSQK[a]	2.2 (2.99[c])	(3.1[c])	-6.9 (-9.2[c])
Myoglobin			
HGTVVLTALGGILK[a]	1.7 (1.63[d])	(3.71[d])	-6.4 (-6.0[d])
HGTVVLTALGGILK[b]	1.6 (1.63[d])	(3.71[d])	-6.7 (-6.0[d])
VEADIAGHGQEVL[a]	1.7 (1.63[d])	(3.71[d])	-6.4 (-6.0[d])
VEADIAGHGQEVL[b]	1.7 (1.63[d])	(3.71[d])	-7.1 (-6.0[d])
Hb			
TPAVHASLDKF[a]	0.7 (0.7[e])	2.43	-8.2 (-8.3[f])
TPAVHASLDKF[b]	0.4 (0.7[e])	2.43	-8.2 (-8.3[f])
GTFATLSELHCDK[a]	0.7 (0.7[e])	2.43	-8.2 (-8.3[f])
GTFATLSELHCDK[b]	0.4 (0.7[e])	2.43	-8.2 (-8.3[f])
Hb - Hp			
TPAVHASLDKF[a]	3.1 (NA)	1.28	-10.5 (NA)
TPAVHASLDKF[b]	2.5 (NA)	1.28	-10.5 (NA)
GTFATLSELHCDK[a]	3.1 (NA)	1.78	-12.0 (NA)
GTFATLSELHCDK[b]	2.7 (NA)	1.78	-12.0 (NA)

[a]5 days exchange. [b]11 days exchange. [c]Value from reference [40]. [d]Value from reference [41]. [e]Values from reference [39] . [f]Value from reference [42]. "NA" indicates value not available.

2.3.3 Analysis of a Non-Two-State Folding System

BCAII, which contains a total of 11 histidine residues in its primary amino acid sequence and is a known non-two-state folding protein,[43] was also analyzed using the protocol described above. Peptides containing 9 of the 11 histidine residue in BCAII were detected in the LC-MS/MS readout, and 5 of these 9 histidine residues (i.e., His-93,

His-95, His-106, His-118 and His-121) were found to be from globally protected regions in BCAII's native three-dimensional structure, based on the histidine H/D exchange behavior of peptides containing these residues (Figure 7). Histidine-containing peptides containing these 5 histidine residues yielded sigmoidal curves with two clear transitions (Figure 7), with $C_{1/2}$ values of 1.4 M and 3 M. These results are consistent with the presence of a folding intermediate that is stabilized in ~2 M GdmCl. The presence of such an intermediate has been suggested in other chemical-denaturant induced equilibrium unfolding/refolding studies using intrinsic fluorescence spectroscopy [43-44]

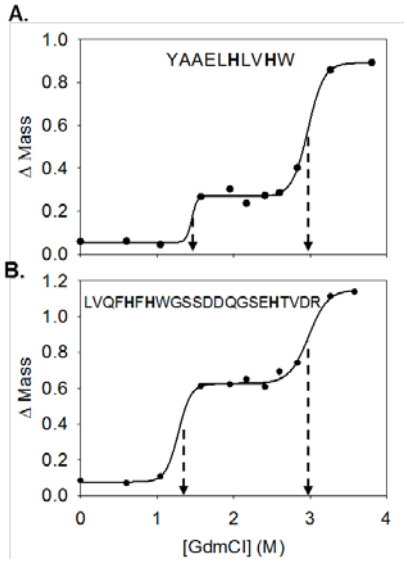

Figure 7: Slow histidine H/D exchange data for BCA II. Data obtained on a peptide containing histidine residues, His-118 and His-121, is shown in (A) and data obtained for a peptide containing histidine residues, His-93, His-95 and His-96 is shown in (B). The dotted arrows indicate $C_{1/2}$ values. The solid lines represent the best fit of the data to equation (13), with the data in each transition being fit separately.

2.3.4 Analysis of a Protein-Protein Interaction

An important application of the histidine H/D exchange protocol developed here is the detection and quantitation of protein-ligand binding interactions. In order to test the ability of the protocol to detect and quantify a protein-ligand binding interaction, the known protein-protein interaction between Hb and Hp was analyzed.[37] A Hb sample and a sample of the Hb-Hp complex were each subject to the histidine H/D exchange protocol described here using H/D exchange times of both 5 and 11 days. In these analyses a total of 9 hemoglobin peptides containing 6 of the 10 histidine residues in the α chain of hemoglobin and 5 of the 9 histidine residues in the β chain of hemoglobin were identified in the LC-MS/MS readout (see Table 2).

The $\Delta Mass_{wt,av}$ values generated for 7 of the 9 histidine-containing peptides in Hb were all similar (i.e., ~0.6 Da) and did not change with denaturant concentration (data not shown) in either the Hb or Hb-Hp analyses, suggesting that the histidine residues in these seven peptides (see Table 2) were solvent exposed in hemoglobin's three-dimensional structure. The $\Delta Mass_{wt,av}$ values recorded for 2 of the 9 histidine-containing peptides detected, Hb (116-126) of sequence TPAVHASLDKF from the α chain and Hb (83-95) of sequence GTFATLSELHCDK from the β chain, showed a denaturant dependence (see Figure 8).

Table 2: Summary of $C_{1/2}$ values determined for the histidine-containing hemoglobin peptides identified in the Hb and Hb-Hp analyses described here. "ND" indicates that no denaturant dependence was observed for the $\Delta Mass_{wt,av}$ values determined for these peptides, presumably because the histidine residues in these peptides were derived from solvent exposed regions of the intact protein structure.

Protein Peptide (amino acid #)	$C_{1/2}$ (M)
Hemoglobin α Chain	
VGAHAGEYGAEALER (17- 31)	ND*
TYFPHFDLSHGSAQVK (41-56)	ND
KVADALTNAVAHVDDMPNAL (61-80)	ND
LVTLAAHLPAEF (106-117)	ND
TPAVHASLDKF (116-126)	0.7(0.4*)
Hemoglobin β Chain	
SDGLAHLDNLK (72-82)	ND
GTFATLSELHCDK (83-95)	0.7(0.4*)
LHVDPENFR (96-105)	ND
VVAGVANALAHKYH (132-146)	ND
Hemoglobin-Haptoglobin α Chain	ND
VGAHAGEYGAEALER (17- 31)	
TYFPHFDLSHGSAQVK (41-56)	ND
KVADALTNAVAHVDDMPNAL (61-80)	ND
LVTLAAHLPAEF (106-117)	ND
TPAVHASLDKF (116-126)	3.1(2.5*)
Hemoglobin-Haptoglobin β Chain	
SDGLAHLDNLK (72-82)	ND
GTFATLSELHCDK (83-95)	3.1(2.7*)
LHVDPENFR (96-105)	ND
VVAGVANALAHKYH (132-146)	ND

* Values after 11 day exchange

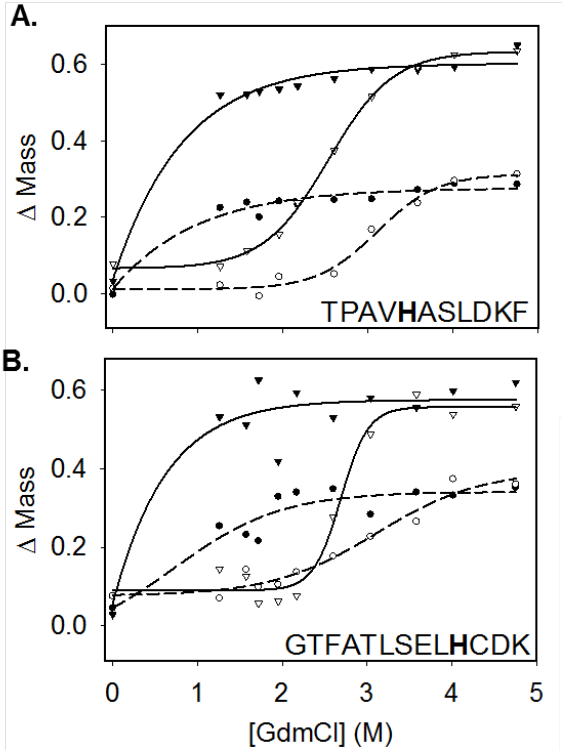

Figure 8: Slow histidine H/D exchange data for Hb and the Hb-Hp complex. Data obtained on His-120 containing peptide from the α chain of Hb after 5 days (○ and ●) and 11 days (▽and ▼) in the presence (○ and ▽) and absence (● and ▼) of Hp is shown in (A). Similar data obtained on a His-92 containing peptide from the β chain of Hb in the presence and absence of Hp is shown in (B). The lines represent best fit of each data set to equation (13).

However, a clear pre-transition baseline was only observed in the analysis of the Hb-Hp complex. Only one data point was collected in the pre-transition baseline of the Hb data sets because the $C_{1/2}$ values obtained for Hb in the absence of Hp were small (i.e., < 1 M GdmCl). The peptides from α and β chains yielded similar $C_{1/2}$ values at similar H/D exchange times (see Table 1). These results suggest that they belong to the same folding unit, even though they were derived from different subunits in the hemoglobin complex. These data are consistent with hemoglobin's three-dimensional

structure, which has the histidine residues in these peptides buried in the heme pocket.[10]

Hb is a tetramer in solution composed of two α chains and two β chains with different amino acid sequences but similar 3-D structures.[10] The GdmCl-induced equilibrium unfolding reaction is known to be biphasic with the Hb tetramer dissociating to two dimers, each containing an α and β chain, before each of the α/β dimers unfold at high denaturant concentration (> 5 M).[39] The $C_{1/2}$ values recorded for the two globally protected histidine-containing peptides in our H/D exchange experiments on Hb are consistent with that expected for the tetramer-dimer transition reported earlier [39]. It is also noteworthy that there is a significant difference (i.e., ~ 0.4-0.6 M) between the $C_{1/2}$ values obtained for the Hb-Hp complex at 5 and 11 days. A similar, but smaller shift of ~0.3 M was observed for Hb at 5 and 11 days for the Hb-Hp complex. As we have previously described for the analysis of SUPREX data, such $C_{1/2}$ value shifts as a function of H/D exchange time can be used to evaluate ΔG_f and *m*-values using the transition midpoint analysis method.[31] The ΔG_f and *m*-values derived for the Hb peptides in the presence and absence of Hp using the transition midpoint analysis method are summarized in Table 1.

The $C_{1/2}$ values obtained for Hb in the presence and absence of Hp indicate that the Hb α/β dimer is stabilized in the Hb-Hp complex. If the ΔG_f values obtained for the two peptides in the Hb analysis are averaged and the ΔG_f values obtained for the two

peptides in the Hb-Hp analysis are averaged, the resulting average ΔG_f values, -8.2 and -11.5 kcal/mol (respectively), can be used to quantify this increased stability (i.e., calculate an average binding free energy of -3.3 kcal/mol). This binding free energy (i.e., $\Delta\Delta G_f$ value) can be used in equation (15) to generate a K_d value of 0.24 nM for the Hb-Hp complex. This K_d value is approximately 10-fold lower than that previously reported using surface plasmon resonance spectroscopy (SPR).[8] The weaker binding affinity measured in the SPR experiment may be a result of the protein immobilization that was necessary in the SPR experiment. It is also possible that the difference may be a result of inaccuracies in our *m*-value assignments, which were obtained by a linear extrapolation involving only 2 data points. Unfortunately, the use of more data points in the linear extrapolation would require the use of impractically long H/D exchange times (e.g., an estimated 3-6 weeks would be required to shift the $C_{1/2}$ values 0.5 M lower than the 11 day $C_{1/2}$ values recorded here) (see Supporting Information).

***m*-value Analysis.** The ΔG_f value calculations described above for Rnase A and myoglobin utilized previously determined *m*-values. In theory, one of the two data analysis methods that we previously described for the evaluation of protein folding *m*-values by SUPREX (i.e., either the transition midpoint analysis method[31] or equation (16)[28]) could be used to evaluate *m*-values in the histidine H/D exchange protocol described here.

$$\Delta Mass = \Delta M_\infty + (\Delta M_o - \Delta M_\infty) e^{\left(\frac{k_\varphi t}{1+e^{-\frac{\Delta G + m[GdmCl]}{RT}}}\right)} \quad (16)$$

In equation (16), ΔM_o is the ΔMass before global histidine exchange, ΔM_∞ is the ΔMass after complete histidine exchange, and the other variables are as described in the text.

In practice, neither of the transition midpoint analysis method or equation (16) could be used effectively in our histidine H/D exchange analyses of Rnase A or myoglobin. Application of the transition midpoint analysis method involves extracting $C_{1/2}$ values from slow histidine H/D exchange data sets obtained using different H/D exchange times and fitting the data to equation (16). However, the extent to which a protein's $C_{1/2}$ value shifts with H/D exchange time depends on the biophysical parameters associated with its folding/unfolding reaction (e.g., ΔG_f and *m*-value). Based on the known ΔG_f and *m*-values of myoglobin and Rnase A, the transition midpoint method would have required the use of impractically long H/D exchange times (several months) to shift the transition midpoints of these proteins by a measurable value (e.g., > 0.3 M).

The ΔMass vs. [GdmCl] data obtained in SUPREX (and in theory the Histidine H/D exchange protocol described here) can be fit to equation (16) to obtain both a ΔG_f and *m*-value. In practice, the accuracy and precision of such ΔG_f and *m*-value determinations is highly dependent on the number of data points that are recorded in the transition region. The lack of sufficient data points in the transition regions of the myoglobin and Rnase A curves (see Figure 5A, 5B and Figure 3A) did not yield reliable

fits of these data sets to equation (16). The ΔG_f calculations in this work on the myoglobin and Rnase A thus employed equation (14) and previously determined *m*-values.

To our knowledge protein folding m-values in GdmCl-induced equilibrium unfolding experiments have not been reported for hemoglobin. The $C_{1/2}$ values recorded for the Hb peptides in this work did shift by a measurable amount using reasonable H/D exchange times of between 5 and 11 days in the histidine H/D exchange protocol. Therefore, it was possible to use the transition midpoint method to evaluate ΔG_f and *m*-values for this protein (see text). We also note that if equation (16) is used to evaluate *m*-values from the Hb-Hp data sets where collected using H/D exchange times of 5 and 11 days (i.e., the two data sets for which 2 or 3 data points were actually obtained in the transition region), an average m-value of 1.9 kcal/(mol M) can be extracted. This value is within 25% of the average value of 1.5 kcal mol^{-1} M^{-1}, which can be determined from the Hb-Hp *m*-values obtained by the transition midpoint method (see Table 1).

2.3.5 Analysis of Proteins in a Yeast Cell Lysate

In order to investigate the scope of the histidine H/D exchange protocol described here, the protocol was applied to the analysis of the proteins in a yeast cell lysate. In this analysis a total of 780 unique peptides from 250 different proteins were identified in the LC-MS/MS analysis, and 93 of the 780 peptides were histidine-containing peptides. It was possible to extract $\Delta Mass_{wt,av}$ values for 50 of these histidine-

containing peptides. In the case of the other 43 histidine-containing peptides it was difficult to extract meaningful ΔMass values as the ion signals for the isotopologues from these peptides were relatively low and/or not well-resolved from other peptides in the mass spectral analyses.

Out of the 50 histidine-containing peptides that were successfully analyzed, 10 histidine-containing peptides from 6 different proteins had $\Delta Mass_{wt,av}$ values with a denaturant dependence. The $C_{1/2}$ values of the peptides ranged from 0.5 to 1.5 M (see Table 3). The results obtained on the four histidine-containing peptides from 3-phospho-glycerate kinase (3PGK) are noteworthy. The three histidine-containing peptides that came from the N-terminal domain of the protein all had a $C_{1/2}$ value of 1.5 M (see Table 3); and the one histidine-containing peptide that came from the C-terminal domain of the protein had a $C_{1/2}$ value of 0.5 M (see Table 3). This is in good agreement with the results of previous protein folding studies on 3PGK that revealed this protein has two different functional domains that fold independently from each other [45].

Table 3: Summary of proteins and peptides that yielded denaturant-dependent histidine H/D exchange behavior

Protein Peptide (amino acid #)	**Transition Midpoint** **$C_{1/2}$ (M)**
3-phosphoglycerate kinase	
AHSSMVGFDLPQR (170 - 182)	1.5
FRHELSSLADVYINDAFGTAHR (148 -169)	1.5
HELSSLADVYINDAFGTAHR (149-169)	1.5
IAGVVYHPSNNELVR (78 - 92)	1.5
ISHVSTGGGASLELLEGK (387 - 404)	0.5
Theoredoxin peroxidase	

HITINDLPVGR (137-147)	1.0
Superoxide Dismutase I	
SVVIHAGQDDLGKGDTEESLK (117-137)	1.3
Enolase II	
NVPLYQHLADLSK (127-139)	1.0
Glyceraldehyde-3-phosphate dehydrogenase	
VINDAFGIEEGLMTTVHSLTATQK (161-184)	1.5
unnamed product gi3724	
VINDAFGIEEGLMTTVHSMTATQK (161-184)	1.0

Superoxide Dismutase 1 (SOD-1) was one of the proteins that yielded a histidine-containing peptide with denaturant dependent $\Delta Mass_{wt,av}$ values in the yeast cell lysate analysis (see Table 3). SOD-1 is a Cu-Zn binding protein. Cu^{2+} is known to strongly affect the stability of protein in vivo but much less is known about the importance of Zn^{2+} on protein stability in vivo [46]. It is noteworthy that the ΔMass versus [GdmCl] plots generated here for SOD-1 yielded a $C_{1/2}$ value similar to that previously reported for the apo form of the protein using a spectroscopic readout [47]. As part of this work we investigated the impact of increasing the Zn^{2+} concentration on the stability of this protein in the context of the whole cell lysate. The results show a $C_{1/2}$ value shift of approximately 0.4 M GdmCl (Figure 9) indicating the SOD-1 was stabilized in the presence of the Zn metal ion in this "*ex vivo*" experiment. In this case the calculation of a K_d value is not possible because a slow histidine H/D analysis of the apo form was not obtained (i.e., the endogenous levels of zinc that were present in the initial cell lysate sample was not known). Nonetheless, our results on SOD can be used to qualitatively identify SOD-1 as a Zn-binding protein.

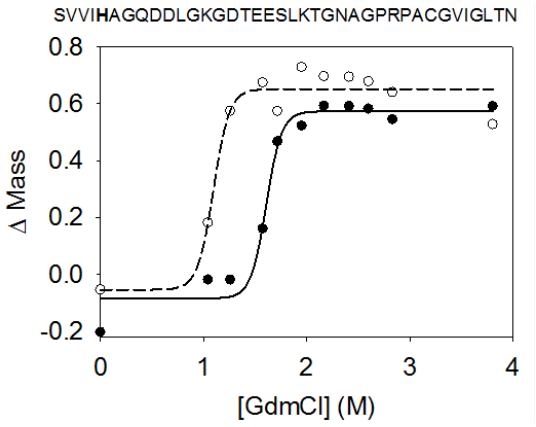

Figure 9: Representative Histidine H/D exchange data obtained from a His-120 containing peptide of superoxide dismutase 1 (SOD-1) in the presence (● and solid line) and absence (○ and dotted line) of added Zn2+.

The 40 histidine-containing peptides that did not show a denaturant dependence to their histidine H/D exchange behavior (see Appendix A) were assumed to be from regions of protein structure that were not globally protected. X-ray crystallographic structures have been solved for 22 of the 40 peptides in Appendix A. A visual inspection of the structures revealed that indeed 16 of these histidine-containing peptides are from solvent exposed regions of protein structure. In the case of 4 peptides, the amino acid sequence region from which they were derived was not included in the solved structure. Only one peptide that did not show a denaturant dependence to its histidine H/D exchange behavior appeared to be from a protected region of structure. However, it is interesting that this peptide from mitochondrial NAD^+-dependent isocitrate dehydrogenase was at a subunit interface in the crystal structure. It is possible that the subunits of this protein may not have been assembled at the protein concentrations used in our experiment.

2.4 Conclusions

The slow histidine H/D exchange protocol outlined here is complementary to other chemical modification and mass spectrometry-based protocols that have been recently described for use in LC-MS/MS based bottom-up proteomics platforms [38, 48] as it provides a new amino acid probe for characterizing the global and subglobal unfolding reactions of proteins and protein-ligand complexes in these experiments. While the results of our cell lysate analysis suggest that a large fraction of the histidine residues in proteins are solvent exposed and not useful for the described protocol, it is noteworthy that many metalloproteins [49] (such as the myoglobin, BCA II, Hb, and SOD-1 proteins analyzed in this work) do indeed have buried histidine residues that are useful for the described protocol. This is because histidine is a common metal ligand in proteins. Thus, the described methodology is likely to be broadly useful for the analysis of ligand binding interactions involving metalloproteins and enzymes.

3. Application of iTRAQ-SPROX protocol to diseased state analysis in Non-Small Cell Lung Cancer

Described in this Chapter is the application of the current iTRAQ-SPROX protocol to the analysis of diseased state in Non-Small Cell Lung cancer. In this study, an iTRAQ-SPROX analysis is performed to compare the thermodynamic stability profile of proteins from a Cyclophilin-A overexpressing lung cancer cell line (ADLC-5M2) and proteins from a Cyclophilin-A knockdown lung cancer cell line (ADLC-5M2-C2).

3.1 Introduction

3.1.1 Cyclophilin A and Lung cancer

Lung cancer is one of the leading causes of cancer death in the US [32]. Despite a lot of efforts in finding biomarkers to lung cancer in particular and to cancer in general, most of the current biomarkers are not predictive [50]. There are a number of cases when biomarkers are just detected at advanced stage [32, 50] and no longer useful to prevent the onset of the disease. Therefore, it is always necessary to discover multiple biomarkers or even a system of biomarkers for more accurate diagnosis and stratification of diseases. A high-throughput approach such as the well-established iTRAQ-SPROX protocol will be well-suited for these studies. The ability to measure thermodynamic stability of hundreds of proteins in one experiment and to quantify the more functional difference between normal vs. diseased state will facilitate the discovery of biomarkers and networks of biomarkers for not only diagnostic but also therapeutic purposes.

The iTRAQ-SPROX protocol is applied for the analysis of two different Lung Cancer Cell lines, one of which has Cyclophilin A overexpressed (ADLC-5M2) and the other has Cyclophilin A knockdown (ADLC-5M2-C2). Cyclophilin A (Cyp-A) is a protein that has been reported to be up-regulated in many different types of cancer such as non-small cell lung cancer [51], breast cancer, pancreatic cancer, colorectal cancer, etc. [52]. However, role of Cyclophilin A cancer is still unclear [52]. To a rough estimation, a Cyp-A overexpressed (Cyp-A (+)) and Cyp-A knockdown (Cyp-A (-)) cell line will resemble the diseased vs. the normal state in which Cyp-A (+) represents the diseased state and Cyp-A (-) represents the normal one. The use of Cyp-A knockdown is more convenient than the use of primary lung cell line for couple reasons. Firstly, this experiment ensures that the majority of the thermodynamic stability difference between the "diseased" and "normal" state will come from the change in Cyp-A expression level. Secondly, the knockdown cell line Cyp-A (-) is also a cancer cell line that doubles almost infinitely as opposed to only 50±10 populations in the primary lung cell line WI-38. This is convenient for the purpose of this preliminary study because of the unlimited resource of biological samples for replicates. A longer term goal of this work is to apply the methodology to more clinical related samples when the protocol is optimized.

3.1.2 The iTRAQ-SPROX protocol

The Isobaric mass tagging in combination with SPROX protocol (see Introduction) has proved to be amenable to large scale analysis of thermodynamic

stability and protein-ligand binding interactions [35]. In a recent study combining SPROX with TMT (Tandem Mass Tag, Thermoscientific), SPROX was able to simultaneously assay 327 proteins in a yeast cell lysate for binding to an immunosuppressive drug Cyclosporine A (CsA) [35a]. This study identified a total of 10 protein targets of CsA including both direct and presumably indirect interaction. A known direct interaction was Cyclophilin A, a specific protein target of CsA with previously determined K_d value of 70nM; the known indirect interaction detected was between CsA and UDP-glucose-4-epimerase. An improvement of the large-scale SPROX methodology by implementing iTRAQ (**i**sobaric **T**ag for **R**elative and **A**bsolute **Q**uantitation) and a Methionine enrichment step was recently reported [35b, c]. This protocol involved the use of a commercially available resin to chemo-selectively isolate the un-oxidized methionine containing peptides in the large-scale SPROX experiment. The protocol was applied to evaluate the interactions of proteins from a yeast cell lysate with a well-known enzyme cofactor β-nicotinamide adenine nucleotide (NAD+) and a less well-understood biologically active ligand, Resveratrol. A total of 232 peptides corresponding to 122 proteins were effectively assayed in the NAD+ binding study and 410 peptides corresponding to 243 proteins were assayed in the Resveratrol binding study. The implementation of a chemo-selection step for Methionine peptides increased the peptide and protein coverage by 1.5 and 2 fold respectively. Also reported in these studies was an estimate of the false positive rate, which was determined to be on the order of 2-4%.

For this is a rather well-established protocol for thermodynamic stability profiling of large protein mixture, the iTRAQ-SPROX protocol was employed for the study of Cyp-A (+) versus Cyp-A (-) reported here.

The underlying hypothesis in conducting this study is that the proteome from diseased cells have different thermodynamic stability profiles compared to those of the proteins in normal cells. This altered thermodynamic stability can be the result of any one of many factors including, for example: (i) a change in protein expression level; (ii) a change in protein-protein or protein-ligand interactions; or (iii) a change in posttranslational modifications (i.e. phosphorylation, glycosylation, etc.). The main difference between the Cyp-A (+) and Cyp-A (-) cell lines used in this study was the expression level of Cyp-A. Overexpression of Cyp-A can introduce changes in protein thermodynamic stability resembling the diseased state, e.g. (i) a change in expression levels of other proteins that are genetically regulated by the abundance of Cyp-A (ii) a change in protein-protein interactions between other proteins and Cyp-A or (iii) a change in posttranslational modifications resulting from downstream effect of Cyp-A overexpression. This study is expected to produce a better molecular level understanding of the functional consequences of Cyp-A overexpression in lung cancer.

3.2 Experimental Procedures

3.2.1 Cell line maintenance

The ADLC 5M2 parental cell line and 5M2-C2 Cyp-A knockdown cell line were originally generated by Howard and coworkers in the laboratory of Dr Edward F. Patz Jr and were given as a kind gift [53]. Both cell lines were maintained in R10+ media that contains the following: RPMI-1640, 10% FBS, 1mM Na-pyruvate, 10mM HEPES, 0.25 (w/v) glucose and 100 Units/mL penicillin and 100 μg/mL streptomycin in a total of 500 mL. The cells in a frozen stock of ADLC 5M2 and ADLC 5M2-C2 were separately inoculated into pre-warmed 10 mL of R10+ media and subsequently passaged to the T-150 flask and grown to almost 100% confluency. Cells were harvested with HQTase, centrifuged at 1,000xg for 10 min at 4°C before being washed with PBS buffer and kept frozen at -20°C until subsequent SPROX analysis.

3.2.2 Western Blot

1 pellet of each cell line was obtained and 500 μL of Lysis buffer (20 mM pH 7.4 and 1 X Halt inhibitor cocktail) was added to each pellet. The cells were lysed by 10 cycles of disruption with 1 mm glass silicon beads (biospec) for 20 sec and incubation on ice for 1 min. Protein concentrations were measured by Bradford assay to be 13.2 mg/mL for ADLC 5M2-C2 and 12.8 mg/mL for ADLC 5M2. The lysates were kept at -20°C overnight before subject to gel electrophoresis. A 5 μL of undiluted Lysate was combined with 95 μL of Sample buffer. Sample buffer contains 1X NuPAGE sample

buffer, 50mM DTT and H_2O. The lysate was subsequently diluted 2, 4 and 8 fold and a 20μL was loaded to the NuPAGE precast polyacrylamide gel for 1-D SDS PAGE. The loading order (from right in Figure 10) was as followed: 8 fold diluted 5M2-C2, 8 fold diluted 5M2, 4 fold diluted 5M2-C2, 4 fold diluted 5M2, 2 fold diluted 5M2-C2, 2 fold diluted 5M2, 1X 5M2-C2 lysate, 1X 5M2 lysate, positive control (5 ng Cyp-A), blank and Marker. The gel was allowed to run for 30 min at 200 V. Proteins were transferred to PVDG membrane for an hour. Resulting transferred proteins were confirmed with Ponceau Dye, wet with Methanol, washed with diH_2O and blocked with 5% milk/PBST (1X PBS, Tween-20 and H_2O). The protein blot was incubated with 1st Antibody (rabbit anti-Cyp-A igG) with 1:1000 fold dilutions to 5% milk/PBST for approximately 1 hr. The protein blot was rinsed with twice PBST before binding to the 2nd Antibody; anti-rabbit igG (1:10,000 fold dilutions). The protein blot was rinsed twice with PBST again before incubation with the West Pico mixture and exposed for 1 sec, 3 sec… 3 min. The result shown in figure 10 is from 1 sec exposure.

3.2.3 SPROX Analysis

The Cyp-A (+) and Cyp-A (-) lysates were prepared from 2 comparable cell pellets each. Each pellet had a number of cells equivalent to a monolayer on 2 T-150 cell culture flasks. The pellets were lysed using the procedure reported in Chapter 2-Histidine HDX. The total protein concentration in each lysate was determined using a Bradford assay. The concentrations were 9 mg/mL and 7.5 mg/mL for the Cyp-A (+) and

Cyp-A (-) lysates respectively. The total protein concentration in each lysate was adjusted to 7 mg/mL.

SPROX analysis was initiated by dilution of 25 μL of each lysate (i.e. Cyp-A (+) and Cyp-A (-)) to a series of eight 20 μL aliquots of denaturant containing buffers. Each of the two series contained the following GdmCl concentrations: 1.2, 2.5, 3.1, 4.3, 5, 6.2, 7.1 and 8.3 M. The lysates were allowed to equilibrate in denaturant containing buffers for approximately 30 min before oxidation was initiated by adding 5 μL of 30% (w/v) H_2O_2 (approximately 10 M). The final GdmCl concentration at the time of the oxidation reaction was: 0.5, 1, 1.3, 1.7, 2.0, 2.5, 2.8, 3.3 M.

Oxidation reaction was allowed to proceed for 3 min, which is 3 half-lives of the oxidation reaction for an un-protected methionine residue, before the reaction was quenched with 1 mL of saturated solution of L-Methionine (300 mM). An aliquot of 100% (w/v) TCA solution was added to each SPROX sample to a final concentration of 10% and incubated overnight to precipitate proteins. The resulting protein pellets were washed with 300 μL Ethanol three times and digested with trypsin; resulting peptide digest were labeled with iTRAQ reagent according to protocol that has been reported elsewhere[35c].

3.2.4 Methionine Enrichment

An 80 μL aliquots of each iTRAQ labeled samples was subject to a Methionine Enrichment step using Pi3-Methionine Resin according to the reported protocol [35c] with

the following exceptions (i) the 80 μL samples were speed-vac down to approximately 30 μL and diluted with 60 μL of Acetic Acid (ii) the peptide solution was incubated with the methionine capturing resin for 2.5 hours instead of 1.5 hours.

3.2.5 LC-MS/MS Analysis and Data Analysis

A 50 μL aliquot of each iTRAQ labeled sample was combined resulting in approximately 400 μL. A 100 μL aliquot of this combined sample was subject to a C18 desalting step, the desalted peptides were eluted to a total of 150 μL of 70% Acetonitrile. These desalted peptide solutions were speed-vac down to dryness, brought back to 10 μL of 100% acetonitrile and diluted with 240 μL of buffer A. A 40 μL aliquot was subject to LC MS/MS analysis on the Agilent QTOF in a similar manner to the previously reported protocol[35c]. A total of 5 LC MS/MS runs were performed for each sample. The Data Analysis is performed in similar manner to previously reported protocol with the following exception: the matched peptides are directly visually inspected for $C_{1/2}$ shift instead of subjecting to difference analysis.

3.3 Results and Discussion

3.3.1 Western Blot data confirms Cyp-A knockdown

Three replicate Western Blot experiments were performed to confirm knockdown of Cyp-A in the Cyp-A (-) cell line. Presented in Figure 10 is representative the Western Blot result from the third replicate. The result shows an approximate 3 fold

decrease of the expression level of Cyp-A in Cyp-A (-) with respect to Cyp-A level in Cyp-A (+) cell line.

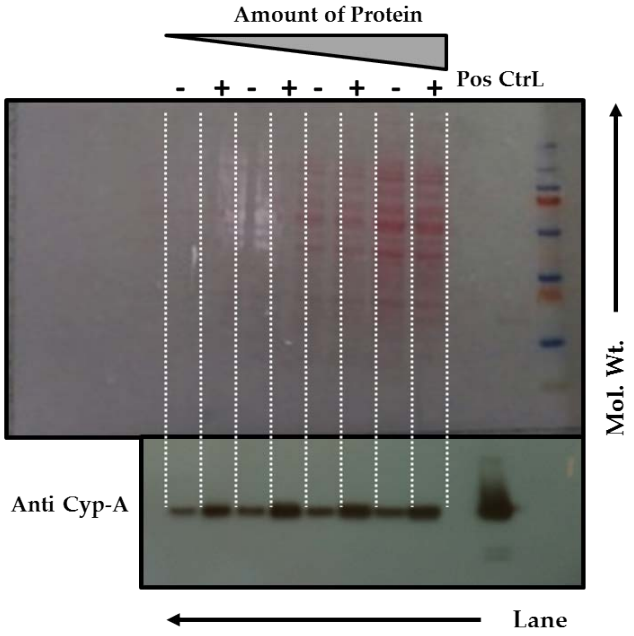

Figure 10: Representative Western Blot Result of proteins from Cyp-A parental (+) and Cyp-A knockdown (-) cell line. Upper panel is an image of the ponceau stained PVDG blot preceding tranfering showing decreasing total protein amount loaded on to each lane. Lower panel is the western blot result showing decreased expression level of Cyp-A in the knockdown Cyp-A (-) cell line.

3.3.2 General Strategy

The iTRAQ-SPROX protocol used in this study is shown in Figure 11. Initially, equal amounts of proteins from Cyp-A (+) and Cyp-A (-) cell lysate were subject to simultaneous SPROX analyses as reported elsewhere [35b, c]. Once the oxidation reaction was quenched, the resulting proteins in each SPROX sample were subject to TCA precipitation, re-dissolvation, reduction, alkylation and protease digestion. The resulting peptide digests were labeled with each of the iTRAQ 8-plex reagents in such a way that

the lowest denaturant concentration corresponded to the lowest isobaric tag (113) and the highest denaturant concentration corresponded to highest isobaric tag (121). After labeling with isobaric mass tags, the peptide samples were combined into a single sample, resulting in one sample from the Cyp-A (+) and one sample from the Cyp-A (-) cell lines.

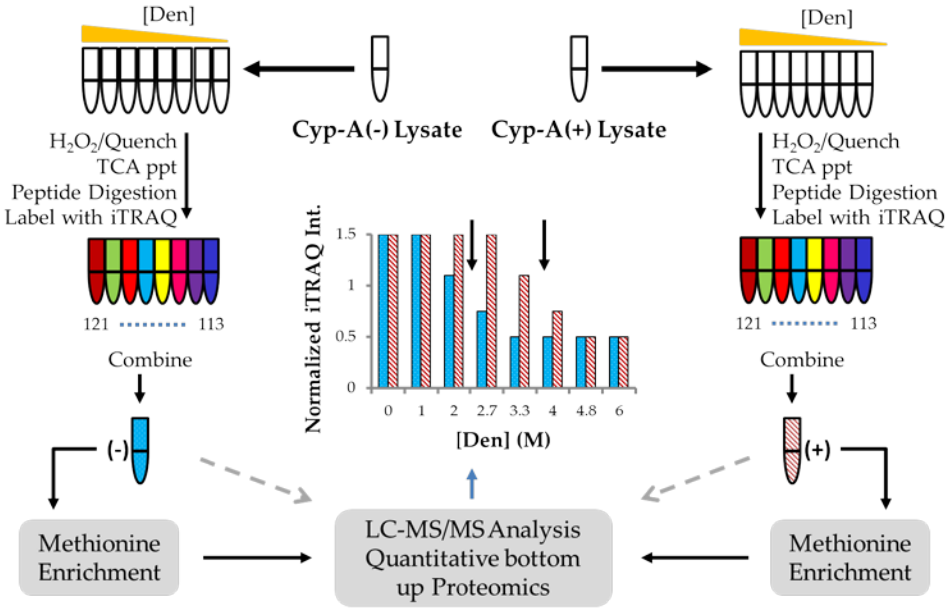

Figure 11: General strategy for the iTRAQ-SPROX protocol

Once combined, the peptide solutions were subject to LC MS/MS analyses either directly or after a Methionine enrichment step. The role of the methionine enrichment step is to chemically select the un-oxidized methionine- allowing the oxidized methionine- and non-methionine- containing peptides to flow through. The data obtained on the direct submission of iTRAQ labeled samples (i.e. without methionine enrichment) is called the control data set. This data contained identification and

quantitation information for (i) non-methionine- (ii) oxidized methionine- and (iii) un-oxidized methionine- containing peptides. The data obtained on the non-methionine-containing peptides are expected to be unaltered by the oxidation reaction and/or denaturant concentration. Therefore any changes observed for the non-methionine-containing peptides can be results from experimental errors (i.e. random iTRAQ quantitation error, differential precipitation and re-dissolvation in the TCA precipitation step, mixing error, etc.). As a result the data from the non-methionine containing peptides in the "control" runs can be used to generate "normalization factors" to correct for the experimental errors may be introduced to sample preparation steps. After applying normalization on methionine- containing peptides from both the control and the methionine- enrichment runs, the peptides from the Cyp-A (+) samples are matched against the Cyp-A (-) samples. The iTRAQ-SPROX data obtained on common peptides in the Cyp-A (+) and Cyp-A (-) samples are compared and peptides with significant $C_{1/2}$ shift (i.e. ≥0.5 M GdmCl) are identified as hits (See Data Analysis section and reference[35c]).

3.3.3 Proteomic Coverage

Presented in Table 4 is a summary of the proteomic coverage for this iTRAQ-SPROX experiment. The results are typical of those generated in shotgun proteomics experiments conducted on the Agilent QTOF platform used in this work. The number of methionine containing peptides in the control runs represents 20% of the total peptide

identifications (IDs) and approximately 30% of the total protein IDs. Successful methionine enrichment added an extra 186 methionine-containing peptides and 82 methionine-containing proteins to the analysis of the Cyp-A (-) samples corresponding to a 3-fold increase in the number of peptides and 2-fold increase in the number of proteins observed. Methionine enrichment of the Cyp-A (+) samples on the other hand only added an extra 86 peptides (1.5 fold the number of peptides) and 3 extra proteins to the analysis. A total of 282 peptides (164 proteins) from the Cyp-A (-) and 216 peptides (114 proteins) from the Cyp-A (+) samples were matched, and the resulting 120 matched peptides (57 proteins) were further analyzed for $C_{1/2}$ shift greater than or equal to 0.5 M GdmCl.

Table 4: Proteomic coverage of the iTRAQ-SPROX Cyp-A (+) vs Cyp-A (-) experiment

Runs	*Criteria*	*CypA (-)*	*CypA (+)*
Control (5 runs)	Peptide identifications	1329	1669
	Unique Peptide IDs (proteins)	464 (304)	543 (345)
	Unique Met-Peptides (proteins)	96 (82) (67 oxidized)	130 (111) (99 oxidized)
Met-enrichment runs	Peptide identifications	385	437
	Unique Peptide IDs (proteins)	148 (96)	157 (108)
	Unique oxidized-Met-peptides (proteins)	29 (21)	32(22)
Matched peptides (proteins)		**120 (57)**	

The methionine-enrichment results from the Cyp-A (-) samples are comparable to that of the recently reported NAD+ binding study[35b]. Due to less successful

methionine enrichment in the Cyp-A (+) samples, the amount of matched peptides (and proteins) are half as much of the NAD+ binding study. Nevertheless, this result is comparable to that of previously reported iTRAQ-SPROX data sets and considered fairly standard outcome.

3.3.4 iTRAQ-SPROX Analysis

Experimental errors that can arise from (i) iTRAQ reporter ion intensity difference from peptides to peptides and (ii) different iTRAQ reporter ion intensity of 1 peptide from tags to tags. Therefore, two normalization steps were performed on the each raw iTRAQ reporter ion intensities of a peptide. The first normalization (N1 normalization) takes into account the difference in iTRAQ reporter ion intensity from tags to tags. That is, all intensity across different reporter ions (i.e. iTRAQ tags) was averaged for 1 particular peptide sequence. This averaged intensity is then used to normalize all intensity from each tag (i.e. by dividing the raw intensity from each tag by the averaged intensity). This N1 normalization accounts for the random error of the iTRAQ reporter ion intensities (from tags to tags). The second normalization factor (N2 normalization factor) was determined by averaging all intensity of different peptides from the same tag (from peptides to peptides). For instance, each peptide sequence contains a specific 113 reporter ion intensity. The averaged value of all N1 normalized 113 reporter ion intensities of all peptides will give the N2 normalization factor for the 113 tag. The variation of the N1 normalized 113 reporter ion intensities is considered the

variation of the iTRAQ quantitation in this experiment. This normalization step accounts for systematic difference of the total peptides amount generated from each SPROX samples (i.e. each denaturant concentration). This difference can arise from TCA precipitation, protease digestion and iTRAQ labeling. The N2 normalization factor is summarized in Table 5.

Table 5: Normalization factors of the 8 iTRAQ reporter ions in the iTRAQ-SPROX experiment

	113	**114**	**115**	**116**	**117**	**118**	**119**	**121**
N2 Factor	0.96	0.88	0.89	1.09	1.05	0.997	0.99	1.03
STDEV	0.47	0.37	0.36	0.41	0.38	0.40	0.397	0.43

As can be seen from table 5, most of the averaged values of the N1 normalized reporter ion intensities (i.e. the N2 normalization factors) centered around 1 representing reproducible sample preparation from samples to samples (i.e. different denaturant concentrations). The standard deviations are relatively similar with range from 36 to 47%, which also represents typical iTRAQ quantitation error in the iTRAQ-SPROX analysis on the Agilent QTOF platform. The N2 normalization factors and standard deviations are not affected by the identification scores, therefore were used to generate N2 normalized iTRAQ intensities for subsequent data analysis.

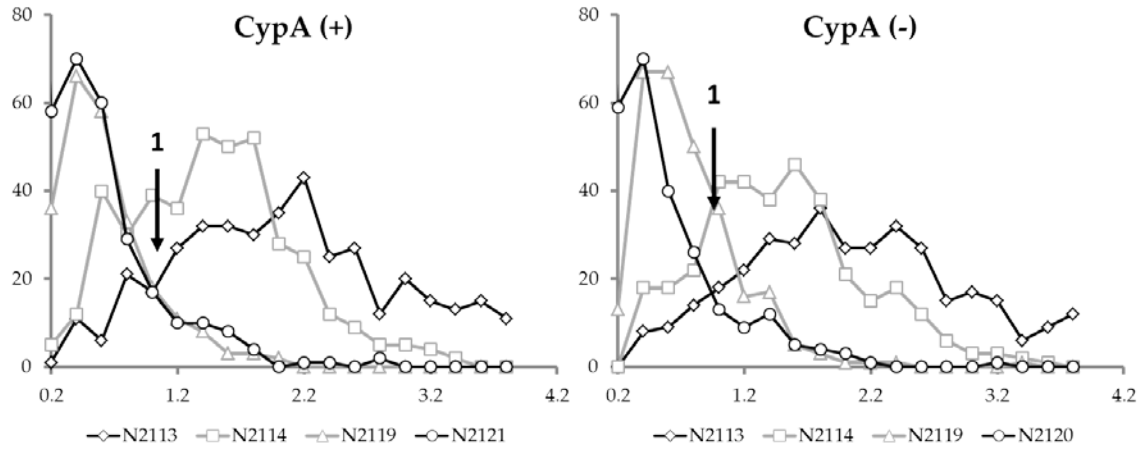

Figure 12: Distribution of the iTRAQ intensities of the 113 and 121 reporter ions for un-oxidized methionine containing peptides from Cyp-A (+) on left and Cyp-A (-) samples on right. Black arrows indicate intersection of the 2 distributions (113 vs. 121). Distribution of the 114 and 119 tags are also included for comparison.

Presented in Figure 12 is the distribution of iTRAQ intensities for all the un-oxidized methionine containing peptides for the lowest isobaric tag (113) and highest isobaric tag (121) representing the lowest and the highest denaturant concentration, respectively. In the SPROX experiment, most proteins remain folded at low denaturant concentration protecting the buried methionine residues from being oxidized by H_2O_2. Increasing denaturant concentration increases the unfolded protein population hence increasing extent of oxidation at the globally protected methionine residues. Therefore the un-oxidized methionine containing peptides will appear predominantly at low denaturant concentration, i.e. the 113 tag will have higher normalized iTRAQ intensity and the 121 tag will have lower normalized iTRAQ intensity. This is indeed the case for un-oxidized methionine containing peptides from both Cyp-A (-) and Cyp-A (+)

samples. The intensities of the 113 tag distribution centered around 0.4 and the intensities of the 121 tag distribution centered around 1.8-2.2. The intersection of the 2 distributions (i.e. 113 tag intensity and 121 tag intensity) is considered the averaged transition midpoint and is utilized herein as a "cut-off" line between the pre- and post-transition baseline. That is, the iTRAQ intensity above this line (of 1) is considered "pre-transition" and below this line is considered "post-transition". Any iTRAQ intensity within 10% of the transition baseline is considered "at the transition mid-point" and the corresponding denaturant concentration is defined as the $C_{1/2}$ value. Using these criteria, a total of 6 of hit proteins were identified with altered thermodynamic stability in the presence and absence of Cyp-A overexpression (see below).

3.4 Hit Proteins Identified

Table 6: Protein hits that show changes in thermodynamic stability in the presence and absence of Cyp-A overexpression.

Protein	*Peptide*	$\Delta C_{1/2}$ **(M)**
Fructose-bisphosphate Aldolase A isoform-1	(K)FSHEEIAMATVTALR(R)	0.8
Phosphoglycerate kinase-1	(K)ALMDEVVK(A)	-0.7
Alpha-Enolase isoform-1	(K)LMIEM(ox)DGTENK(S)	0.7
Argininosuccinate synthase	(R)MPEFYNR(F)	0.6
Beta-tubulin	(K)MAVTFIGNSTAIQELFK(R)	0.9
Eukaryotic Translation Initiation Factor 5A-2	(K)IVEMSTSK(T)	0.5

Summarized in Table 6 are the proteins (and peptide) hits identified in this iTRAQ-SPROX experiment and their corresponding $C_{1/2}$ shifts (all are significant shifts ≥

0.5 M GdmCl). Interestingly, β-tubulin was identified to be overexpressed in non-small cell lung cancer by Okuzawa and coworkers using 2-Dimensional gel electrophoresis and nonenzymatic sample preparation[54]. Chen and coworkers also reported the overexpression of eIF-5A in lung cancer using 2-DE analysis with identification using mass spectrometry and 2-Dimensional Immunoblots [55]. Both of these proteins are stabilized by overexpression of Cyp-A in the parental cell lines ADLC-5M2 suggesting possible interactions of these proteins with Cyp-A in the development of lung cancer. Presented in Figure 13 is representative iTRAQ-SPROX data from β-tubulin and Eukaryotic Translation Initiation Factor 5A-2.

It is important to note that the false discovery rate of a typical iTRAQ-SPROX experiment can be as high as 5%[35b] which make the possible number of falsely identified peptide hits in this assay to be ~6 (i.e,5% of 120 times), which is equivalent to the to the number of hits identified. Also, it can be seen from Table 6 that all proteins are identified with 1 peptide hit. Thus, it is possible that these peptide hits are falsely discovered, and they do not have any biological significance. However, based on the results of previous biomarker studies in lung cancer, there may be some real biological significance to at least some of these hits.

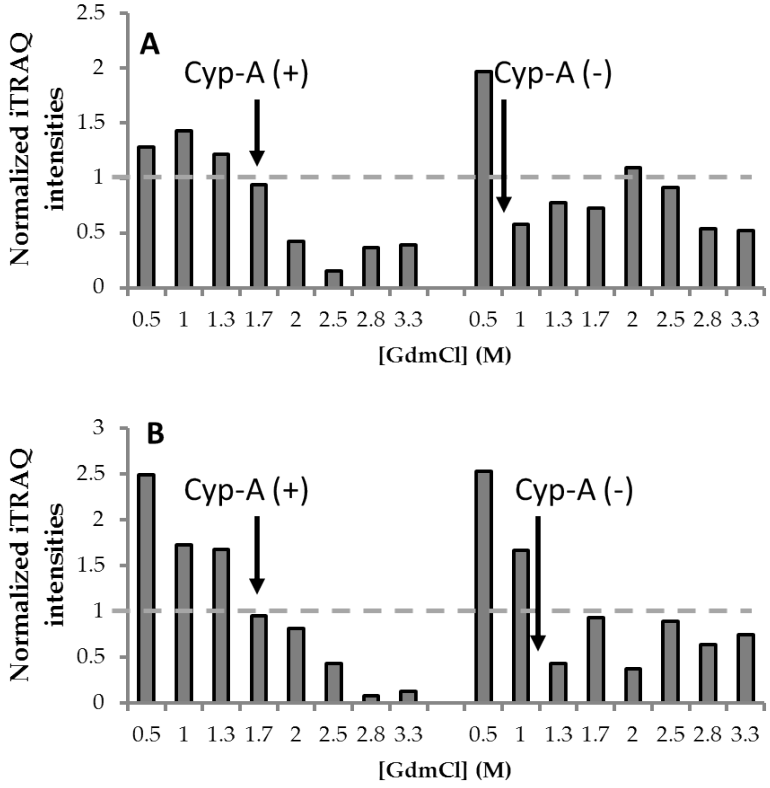

Figure 13: Representative iTRAQ-SPROX results from β-tubulin (A) and iEF5A (B). Bar graphs on the left represent peptides generated in Cyp-A(+) sample and bar graphs on the right represent peptides generated in Cyp-A (-) sample. Black arrow indicates estimated $C_{1/2}$ value and dotted line represents the "cut-off" line (see text)

3.5 Conclusions

The iTRAQ-SPROX protocol presented here has been applied to the thermodynamic stability differentiation analyses of the proteins from a Cyp-A overexpressing lung cancer cell line versus proteins from a Cyp-A knockdown lung cancer cell line. The protocol has assayed a total of 57 proteins and identified 6 potential protein hits whose thermodynamic stability changes in the presence and absence of Cyp-A overexpression; two of which are known lung cancer biomarker discovered in

previous proteomics studies (i.e. β-tubulin and eIF5A). More biological replicates are needed to confirm the reproducibility of these potential hits. The current iTRAQ-SPROX protocol utilized here is amenable to this type diseased state of analyses, however the proteome coverage is largely limited to the identified and quantified methionine containing peptides in the bottom-up proteomics readout. To obtain more comprehensive analyses in the human proteome such that in the Lung cancer cell lines, proteomics coverage of this experiment needs to be improved.

4. Development of a SILAC-SPROX protocol and application to ATP binding discovery

4.1 Introduction

4.1.1 Motivation

As described in previous chapter, current iTRAQ-SPROX methodology has several drawbacks. One drawback is that in the iTRAQ-SPROX protocol (see Figure 11), the (+) and (-) ligand samples at different denaturant concentrations are prepared separately and are not labeled with iTRAQ reagents until the end of the protocol. This means that technical errors associated with various sample handling steps, e.g. TCA precipitation, re-dissolvation and protease digestion, can arise. The inherent technical error associated with iTRAQ quantitation (i.e., the error associated with the LC MS/MS readout) is approximately ± 10%[56]. However, due to the aforementioned issues related to sample handling in the SPROX experiment; the variation in iTRAQ-SPROX experiments can reach 30-40%. This error is acceptable for measurement of large changes in iTRAQ intensities. However, a typical iTRAQ-SPROX curve has relatively few points (i.e., 8 points) and relatively small amplitude (~1 normalized iTRAQ intensity unit). This ultimately makes the $C_{1/2}$ assignment challenging.

Another drawback is that iTRAQ quantitation is based on the data collected in product ion mass spectra, in which complications can arise from the analysis of chimeric peptides in which two precursor ions with similar m/z are simultaneously subject to CID

and sequenced. This has been demonstrated to be a significant source of error in the isobaric mass tag quantitation strategy[57].

A third and major drawback to the iTRAQ-SPROX methodology developed to date is that it requires the detection and quantitation of methionine-containing peptides to report on the thermodynamic stability of the proteins to which they map. While the frequency of methionine residues in proteins is relatively low (~2.5%), the large majority of proteins have at least 1 methionine[58]. Thus, the scope of SPROX is not fundamentally limited by the relatively low frequency of methionine residues in proteins. Rather, the protein coverage in proteome-wide SPROX experiments is limited by the practical considerations associated with the comprehensive detection of methionine-containing peptides in the bottom-up shotgun proteomics experiment. In the iTRAQ-SPROX experiment quantitative data is only generated in product ion mass spectra. Therefore, if a methionine containing peptide is not selected for fragmentation (i.e., a product ion mass spectra is not collected), the quantitation information for that peptide will be missing.

The stable isotope labeling with amino acids in cell culture (SILAC) approach has been widely used in mass spectrometry-based proteomic studies of gene expression levels[19]. Proteome-based expression profiling studies using SILAC and other quantitative proteomics technologies are now commonly used to characterize drug-mode-of-action and to understand the basic physiological processes and biological

pathways involved with aging and disease[26]. The use of SILAC quantitation in SPROX experiments has the potential to overcome some of the drawbacks of iTRAQ quantitation described above.

Firstly, the (+) and (-) ligand samples are labeled by amino acid in cell culture at the protein level allowing combining the (+) and (-) ligand samples with matching denaturant concentration right after the SPROX reaction and before proteomics sample preparation. Ultimately, the ratio between (+) and (-) ligand samples are measured within each denaturant concentration, allowing direct comparison of the extent of oxidation of proteins in presence and absence of ligand. This means that the (+) and (-) ligand samples are prepared simultaneously for each denaturant concentration, reducing errors from sample handling steps. Secondly, SILAC quantitation relies on data collected in the first stage of mass spectrometry in LC MS/MS experiments, so, the problem with chimeric peptides in product ion mass spectra (i.e., the second stage of mass spectrometry in the LC-MS/MS experiment) should only affect identification but not quantitation of a certain peptide. A third advantage of SILAC-SPROX experiments is that since SILAC is based on MS1 quantitation, a methionine containing peptide must only be identified in at least on LC-MS/MS analysis for successful analysis. For methionine peptides that ionize but are not actually selected for sequencing in all samples, the identification can be translated based on m/z and retention time from previous LC MS/MS analysis. This allows measurement of (+) to (-) ligand ratio even

when the peptide is not actually sequenced. Lastly, a major advantage of the SILAC-SPROX protocol is that it can be used in combination with a cyanogen bromide cleavage reaction and gel-based fractionation protocol to significantly expand the scope of SPROX experiments, as described in Chapter 5.

4.1.2 Discovery of ATP-binding proteins

To access the validity, scope and sensitivity of the SILAC-SPROX protocol described in this chapter, an ATP-binding experiment was performed using proteins from a yeast cell lysate. ATP (Adenosine-5′-triphosphate) is a common enzyme cofactor that binds to a variety of different protein classes with a wide range of binding affinities. However little is known about binding properties of ATP to proteins in large scale. Two large scale analyses of ATP-binding have been reported, to date. One ATP-binding study involved the use of an energetics based approached called "pulse proteolysis" to detect ATP-binding in proteins from an *E. coli* cell lysate [21a, b]. The other involved the use of an active site reactive immobilized ATP probe (desthiobiotin-conjugated ATP) and affinity pull-down strategy to identify ATP-binding in proteins from a Mycobacterium tuberculosis cell lysate [59].

In the pulsed-proteolysis study, two main strategies were employed to fractionate and simplify the *E. coli* proteome. The first strategy, which involved the use of 2-D SDS-PAGE, claimed to observe approximately 500 proteins and identified 10 ATP-binding targets [21a]. 7 out of 10 ATP binding hits found in this experiment were

previously annotated with ATP-binding GO-term (GO:0005524) from the EcoCyc database. The three ATP-binding hits that were not previously annotated with GO-term ATP-binding were: dihydrolipoamide dehydrogenase (E3), Glyceraldehyde 3-phosphate dehydrogenase A (GAPDH), and periplasmic-binding protein (mlaC). Subsequent validating experiments revealed 1 possible false positive hit (mlaC), which brings the false positive rate to approximately 10%. The second "simplified" strategy involved up-front fractionation of the *E. coli* proteome by ion exchange chromatography and pulse proteolysis incorporated with 1-D SDS-PAGE[21b]. In this study, a total of 30 ATP-binding hits were identified, 21 of which were known to bind ATP (70%). A total of 6 out of the 9 remaining proteins (66%) were not known to bind ATP, but do bind ligands that are similar to ATP (e.g., the nucleoside (uridine), the nucleotide (GTP), or cofactors with adenosine moiety (NAD^+ and FAD)). One of the previously unknown ATP binding proteins identified in this *E. Coli* study, phosphoglycerate mutase (PGM), was further validated in additional experiments.

In the Mycobacterium tuberculosis cell lysate study, a total of 176 ATP-binding proteins were identified, 122 of which (also approximately 70%) are associated with 13 protein families that were annotated with the GO-term ATP-binding[59]. The remaining 54 proteins are presumably proteins that are associated with the ATP-binding hits. Interestingly, a GO-term analysis of the 176 proteins identified to bind ATP reveals

diverse functions associated with small molecule binding (34%), with transferase activity (17%), with oxidoreductase activity (16%), and with DNA binding (9%), etc.

In both of the above studies, there were no estimations of K_d values. However it was noted in the pulse proteolysis study that based on literature search of previously determined K_d from 5 known ATP-binding binding proteins, K_d values are from 0.025 μM to 600 μM. The global analysis of ATP-binding proteins in yeast proteins has not been previously reported. The availability of ATP-binding GO-term and the previous analysis of selected ATP-binding proteins in other organisms help validate the results reported in this chapter and create a means to evaluate the validity, scope and sensitivity of the assay.

4.2 Experimental Procedures

4.2.1 Yeast Cell Lysate preparation

A glycerol stock of the yeast deletion strain BY4739 (Open Biosystems, Lafayette, CO), an auxotroph for lysine, was streaked on a petri-dish containing synthetic complete (SC) media and 30 mg/mL of Light L-Lysine. The SC-media was comprised of 1.7 g of Yeast Nitrogen Base, 5 g of Ammonium Sulfate, 20 g of Bacto-agar, 2 g Glucose, 1.92 g of Synthetic Drop out mix without Lysine (Sunrise Science Product, San Diego, CA). After a 3 day incubation at 30°C, one colony was selected and inoculated into 10 mL of SC-media containing 30 mg/mL of Light L-lysine (Sigma Aldrich) before an overnight incubation at 30°C. A 5 μL portion of this cell culture was transferred into 50 mL (or 100

mL) of SC-media containing Light L-lysine (Sigma Aldrich) at a final concentration of 30mg/mL; while another 5 μL portion of the culture was transferred into 50 mL (or 100 mL) of SC-media containing same concentration of Heavy Lysine ($^{13}C_6$ $^{15}N_2$.Cl). The two cell culture were incubated at 30°C overnight, and a 5 μL portion of each culture was transferred to 50 mL (or 100 mL) of corresponding Light and Heavy media twice before they were ultimately transferred to 1 L of the corresponding Light and Heavy media and grown overnight at 30°C. Portions of the 1 L overnight culture, 60-80 mL each, were centrifuged at 1,000 g at 4 °C for 10 min to generate about 16 cell pellets of the light lysine labeled yeast cells and 12 pellets of the heavy lysine labeled yeast cells.

The Light and Heavy lysine labeled cell lysates (referred to as Light and Heavy lysates) were prepared by combining 400 to 600 μL of lysis buffer to each of cell pellets depending on their size in order to have final protein concentration in the range of 10-20 mg/mL. Lysis buffer was composed of 0.1M Tris•HCl (pH 7.4) and a protease inhibitor cocktail containing 1 mM AEBSF, 500 μM Bestatin, 15mM E-64, 20 μM Leupeptin and 10 μM Pepstatin A. The cells were lysed by mechanical disruption using a disruptor genie (scientific industries) and 400-600 μm diameter glass beads. A total of 15 cycles consisting of mechanical disruption for 30 s and cooling on ice for 1 min were used to lyse the cells in each cell pellet. Ultimately, the lysed cells were centrifuged at 14,000 x g for 10 min at 4 °C, and the supernatant was collected for use in SILAC-SPROX analyses. The concentrations of proteins in the Light and Heavy lysate were determined by

Bradford assay. If the protein concentrations were unequal between the Light and Heavy lysate, both protein samples were normalized to the lower concentration before proceeding to the SPROX analysis.

4.2.2 SILAC-SPROX Analysis

In Solution based experiment 1A/B, 1 volume of 100 mM $MgCl_2$ was added to 8 volumes of both the Light and Heavy lysate; 1 volume of aqueous 828 mM AMP-PNP was subsequently added to the Heavy lysate and 1 volume of Lysis buffer was added to the Light lysate in order to generate the (+) and (-) ligand samples (respectively). Both (+) and (-) ligand samples (i.e. Heavy and Light lysates) were incubated on ice for 30 minutes before a 20 μl aliquot of each sample was combined with 75 μl of each denaturant-containing buffer stock solution, which was prepared in 0.1 M Tris•HCl buffer (pH 7.4) with urea concentrations that ranged from 0 to 8 M. The exact concentration of urea in each denaturant-containing stock solution was determined from a refractive index measurement of the buffer as described previously [60]. The final urea concentration in the 8 denaturant-containing buffers used for these solution based SILAC-SPROX experiments were 0, 1, 2, 2.7, 3.3, 4, 4.8, and 6 M. The samples were equilibrated for 30 min in the denaturant containing buffers before initiating the methionine oxidation reaction in SPROX, which involved the addition of 5 μl of 9.8 M hydrogen peroxide to the protein and protein-ligand samples in each denaturant-containing buffer. The oxidation reactions were each allowed to proceed for 6 min before

they were quenched with 760 μl of a 375 mM solution of L-methionine. The (+) and (-) ligand samples (i.e. Heavy and Light lysate) from the same denaturant containing buffers were combined and 191 μL of 100% TCA (w/v) was added to each combined sample to precipitate the proteins. The samples were incubated overnight at 4°C, and centrifuged at 8,000 x g for 30 minutes at 4 °C, the resulting protein pellets were washed 3 times with 1 mL of Ethanol and 2 times with Acetone and air-dried for 5 min in chemical fume hood before subjected to conventional bottom up proteomics in solution.

The Solution-based ATP-binding experiment 2 was performed as described above for experiment 1 with the following exceptions: (i) the initial $MgCl_2$ concentration was 1M instead of 100mM, (ii) the (-) and (+) ligand samples were incubated on ice for 1 hour and in denaturant containing buffers for 1 hour instead of 30 and 30 min respectively, (iii) the protein samples were distributed into 10 different denaturant containing buffers instead of 8 different buffers with Urea concentration ranged from 0 to 8M; the final Urea concentration in each buffers was: 0.0, 1.0, 1.5, 2.0, 2.5, 2.8, 3.3, 4, 4.9, and 6.0 M, and (iv) the oxidation reaction was allowed to proceed for 24 min instead of 6 min.

4.2.3 Proteomics Sample Preparation

The combined (-) and (+) ligand samples (i.e. Light and Heavy lysates) were subjected to a conventional proteomic analysis. In this proteomic analysis the dried pellets were re-dissolved in 60 μL of 0.1M Tris buffer, pH 8.0, containing 8 M urea;

heated at 37 °C for 5 min; vortexed for 10 min; and heated again at 37 °C for 5 min. A 6 μL aliquot of 50 mM TCEP●HCl and 3 μL aliquot of 2% SDS solution were added to each re-dissolved pellet and the samples were incubated for 1 hr. at 37 °C before 6 μL of a 100 mM solution of MMTS was added to each protein sample. After 10 minute incubation at room temperature, approximately 99 μL of Tris buffer (pH 8.0) containing 4 μg of Lys-C was added to each sample and the proteins were digested overnight at 37 °C. Each sample was centrifuged at 15,000 g for 5 min to precipitate undigested protein. The supernatant from the protein digest was desalted with C18 resin and ultimately eluted in 150 μL of 70% Acetonitrile in 0.1% Formic Acid. A 50 μL aliquot of each peptide solution was dried in a speed-vac (accounted for 1/3 of the cleaned samples), re-dissolved in 50 μL of Buffer A (0.1% Formic acid) and approximately 2 μL of each sample, which contained an estimated 2 μg of peptide, was subjected to LC-MS/MS analyses.

4.2.4 LC MS/MS Analyses

LC-MS/MS analyses were performed using an LTQ Orbitrap XL mass spectrometer (Thermo-Scientific, Inc.). The LC system was configured in a vented format[61] and consisted of a fused-silica nanospray needle packed in-house with Magic C18 AQ 100A reverse-phase media (Michrom Bioresources Inc.) (25cm) and a trapping column containing Magic C18 AQ 200A reverse-phase media (2cm). The peptide samples were loaded onto the column and chromatographic separation was performed

using a two-mobile-phase solvent system consisting of 0.1% formic acid in water (A) and 0.1% acetic acid in acetonitrile (B). The mass spectrometer operated in a data-dependent MS/MS mode over the *m*/*z* range of 400-1800. For each cycle, the five most abundant ions from each MS scan were selected for MS/MS analysis using 35% normalized collision energy. Selected ions were dynamically excluded for 45 seconds.

Raw MS/MS data were submitted to the Maxquant data analysis platform[62] and searched using the Yeast Proteome Database downloaded on 01/09/2013 from ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/proteomes/. The following modifications were considered: methyl methanethiosulfonate (MMTS) at cysteine as a fixed modification; SILAC labeling of lysine (+8), and oxidation of methionine and deamidation of asparagine and glutamine (N and Q). The enzyme was set to Lys-C, and up to 2 missed cleavages were permitted. The false discovery rate was set to 1%. Ultimately, the search output containing the modified peptide sequence, protein name, normalized H/L ratios, score, retention time and raw file name was exported to Excel for further data analysis.

4.2.5 Data Analysis

The protein and peptide identifications obtained from the LC-MS/MS data searches were filtered to contain only H/L ratios ≥ 0. In cases where multiple H/L ratios were obtained for a given peptide sequence and charge state at a specific denaturant concentration (e.g., both the Light and Heavy version of a peptide and charge state were

identified in the LC-MS/MS analysis; and/or a given peptide sequence with a particular charge state was identified in multiple scans); the multiple H/L ratios were averaged to give a single L/H ratio for the peptide with a particular charge state at that specific denaturant concentration. Ultimately, these average H/L ratios were used to generate SILAC-SPROX data sets. Depending on the quality of the data (i.e. the reproducibility of LC-MS/MS runs) and the number of samples with different denaturant containing buffers, a threshold of 4-8 times was used to filter for peptides that appear reproducibly across different denaturant concentration. For instance, the data in the Solution-based ATP-binding experiment 1A was obtained on the Agilent QTOF instrument was filtered to contain peptides identified in 4 or more denaturant concentrations. Whereas the data in the Solution-based ATP-binding experiment 1B, which was obtained on the faster and more sensitive Orbitrap instrument, was filtered to contain peptides appearing in 6 or more denaturant concentrations. The data in the solution based ATP-binding experiment 2 was obtained on the Orbitrap and has 2 more denaturant points was hence filtered to contain peptides appearing in 8 or more denaturant concentrations.

A peptide "hit" was defined as a peptide that has two or more H/L ratios differing by ≥ 1.7 fold with respect to the minimum H/L across the 4 (or 6, 8) denaturant concentrations. An excel formula was used to automatically filter the data sets for peptides that meet those criteria. Ultimately, this list of filtered "hits" was visually

inspected to confirm that their plots of SILAC-SPROX H/L ratios versus denaturant concentrations resemble the structure of the SILAC-SPROX curves depicted in Figure 14.

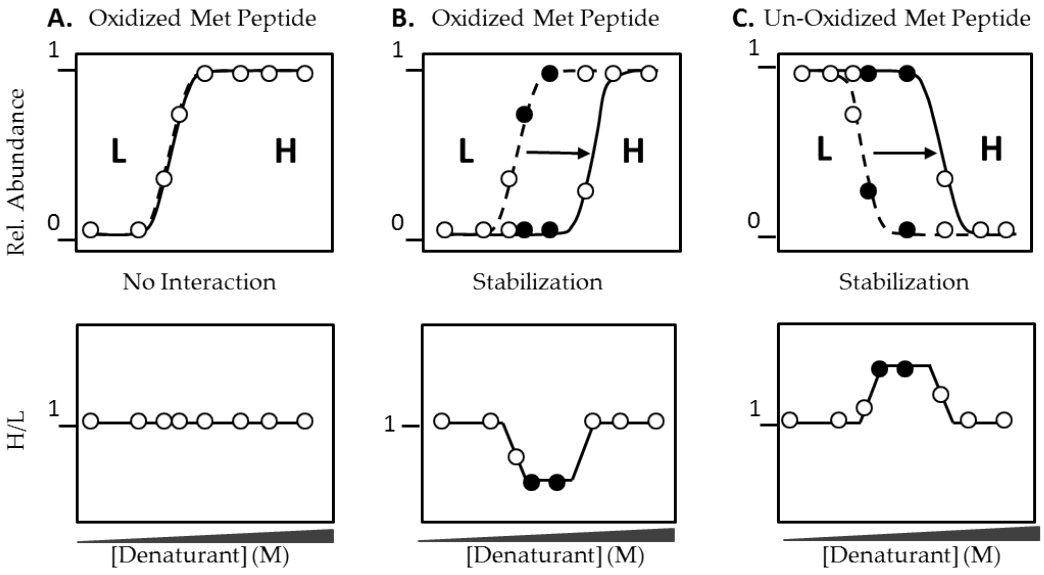

Figure 14: Expected results from SILAC-SPROX solution-based experiments. (A) is an oxidized methionine peptide from a protein that has no interaction with the ligand; (B) is an oxidized methionine peptide from a protein that is stabilized by binding to the ligand; and (C) is a corresponding un-oxidized peptide of that stabilized protein. Open circles represent data points (denaturant concentrations) that have no change in H/L ratio, closed circles represent data points that have significant H/L ratio difference.

4.3 Results and Discussion

4.3.1 General Strategy

The SILAC-SPROX protocol developed in this work is outlined in Figure 15. Initially, "Light" and "Heavy" cell lysates are prepared and the test ligand is spiked into one of the cell lysates. In this case, the ATP analog was spiked into the Heavy lysate. The

two cell lysate samples are subjected to simultaneous SPROX analyses in which aliquots of each cell lysate are distributed into a series of denaturant-containing buffers and reacted with hydrogen peroxide under conditions that we have previously established for the selective oxidation of exposed methionine residues in SPROX analyses (e.g., 0.5 M H_2O_2 for 6 min).[37] The oxidation reactions are quenched (e.g., with the addition of excess L-methionine) and the appropriate Light and Heavy samples with matching denaturant concentrations are combined (see Step 4 in Figure 15). At this point the protein samples can be directly submitted to a conventional bottom-up proteomics analysis in solution (hereafter referred to as the "solution-based" approach).

The SILAC-SPROX strategy described here is analogous to the PrSUIT experiment that has previously been reported[63]. The PrSUIT experiment relied on the use of heavy and light H_2O_2 (i.e., $H_2{}^{18}O_2$ and $H_2{}^{16}O_2$) to define the (-) and (+) ligand samples in SPROX. One disadvantage to PrSUIT is that the differential labeling only incorporates a small mass shift, which complicates the data analysis. The SILAC-SPROX experiment described here has the advantages that the light and heavy samples are separated by at least 8 mass units and the LC-MS/MS data is easily analyzed using well-established SILAC methods. The PrSUIT experiment is also not amenable to the gel-based readout in SILAC-SPROX that ultimately enables the potential binding properties of a protein to be determined using the L/H ratios obtained from any peptides.

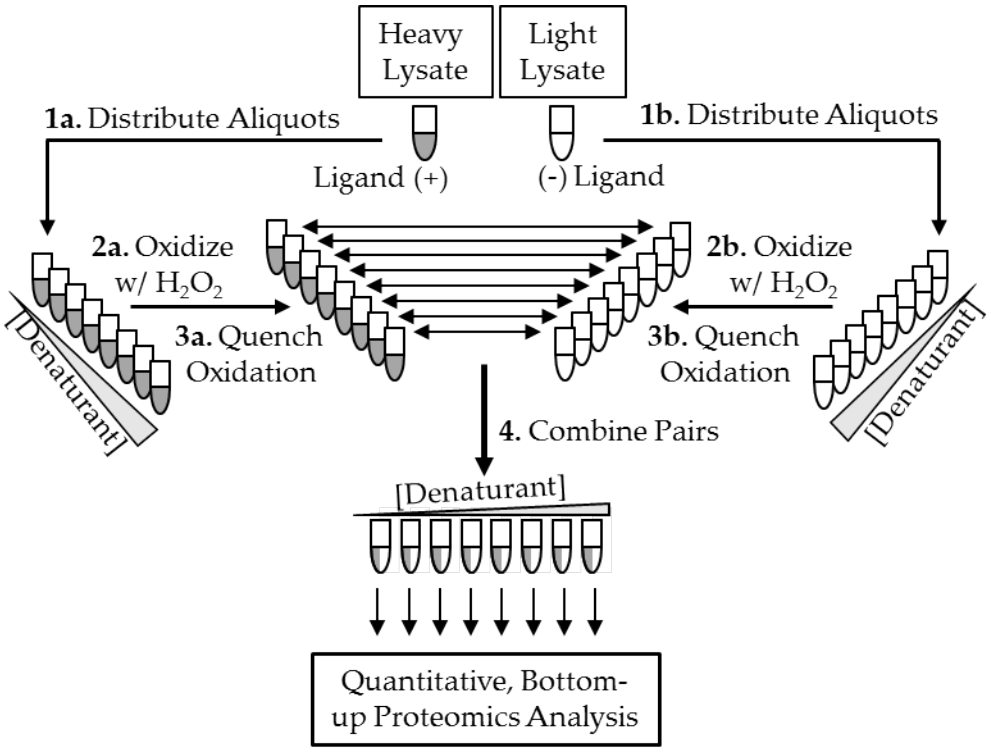

Figure 15: The solution-based SILAC-SPROX protocol

4.3.2 ATP-binding Result Summary

4.3.2.1 Proteome Coverage

The ATP binding properties of the proteins in a yeast cell lysate were analyzed using the SILAC-SPROX protocol developed here (Figure 15). The ligand in the ATP binding study was adenosine 5′-(β,γ-imido)triphosphate (AMP-PNP), a non-hydrolysable ATP mimic that is spiked into the "Heavy" lysate. A total of 2 Solution-based ATP-binding experiments were performed. Other important experimental

parameters (e.g., ATP concentration, equilibration time, oxidation reaction time, and MS instrument platform) used in these experiments are highlighted Table 7.

Table 7: Experiment parameters utilized in ATP-binding solution-based experiments 1A/B and 2

Experiment Name	*MS platform*	*[Ligand]*	*Equilibration Time (min)*	*Oxidation Time (min)*	*Number of Denaturant concentration*
Solution based Exp1A	QTOF	1mM	30 and 30	6	8
Solution based Exp1B	Orbitrap	1mM	30 and 30	6	8
Solution based Exp2	Orbitrap	7mM	60 and 60	24	10

The main difference between experiments 1A/B and 2 was the effective ligand concentrations, equilibration time and oxidation time. As explained in Chapter 1, section 1.4, the sensitivity of this protocol can be tuned by (i) increasing free Ligand concentrations [L] (ii) increasing the incubation time between proteins and ligand for more complete equilibrium and (iii) increasing oxidation time to maximize the linear dependence of $C_{1/2}$ values to changes in folding free energy of the proteins upon binding to ligands. Thus, all the changes made in the solution-based experiment 2 were expected to make the ATP-binding assay in this experiment more sensitive.

It is important to note that the effective ligand concentration is the stoichiometric concentration of the Mg-ATP complex, which is 1 mM in the solution based experiments 1A and B and 7mM in solution based experiment 2. SILAC-SPROX samples from

solution based experiment 1 A and B were subjected to mass spectral analyses using both QTOF and Orbitrap, platforms, respectively, in order to compare the proteome coverage in the two platforms.

Table 8: Proteome coverage and potential protein hits from ATP-binding solution-based experiments 1A/B and 2.

Experiment	*Total Peptides (Proteins) Assayed for Binding*	*Hit Peptides (Proteins)*	*Known ATP Binding Proteins Assayed (Hits)*
Solution-Based Exp. 1A	93 (38)	5 (3)	6 (0)
Solution-Based Exp. 1B	526 (209)	55 (27)	56 (4)
Solution-Based Exp. 2	353 (216)	138 (99)	61 (29)
Total	689 (302)	180 (112)	82 (33)

Summarized in Table 8 are the proteomic results obtained for the above solution-based ATP-binding experiments. On average, when using the Orbitrap mass spectrometry platform, approximately 200 proteins are assayed. The protein included in the assay were those that had at least 1 methionine containing peptide being successfully identified and quantified in at least 6 (or 8) denaturant concentrations. On the other hand, the same samples subjected to LC-MS/MS analyses on the Q-TOF platform resulted in only 38 proteins assayed. These are proteins that also have at least 1 methionine containing peptide being successfully identified and quantified in at least 4 or more denaturant concentrations. This result suggests the use of Orbitrap mass

spectrometry platform gives approximately 5 fold increase in proteomic coverage with respect to the Agilent QTOF platform utilized in these experiments. In total there are 689 peptides corresponding to 302 proteins are assayed across all the solution-based experiments.

The proteomic coverage obtained on the QTOF platform is slightly lower than that obtained on the iTRAQ-SPROX data set reported recently in which approximately 103 peptides corresponding to 70 proteins are effectively assayed[35b]. This is likely due to multiple injections implemented in the LC MS/MS Analyses of the iTRAQ-SPROX samples as oppose to single injection in the SILAC-SPROX samples. When these iTRAQ-SPROX samples were subject to a methionine enrichment step, the proteome coverage increased by approximately 2 fold to 122 proteins assayed. A similar methionine enrichment strategy could be used in SILAC-SPROX; however, it was not done in this work due to the expense associated with performing 8-10 different methionine enrichments on the SILAC-SPROX generated samples.

Presented in Figure 16 are the distributions of the H/L ratios measured for the lysine- and non-methionine-containing peptides identified in the solution-based ATP binding experiment 1B and 2. As shown below in Figure 16, the majority of $\log_2$ (normalized H/L ratio) values in both experiments lie within the range of -0.5 to 0.5. As expected the $\log_2$ values also center around 0 indicating that the large majority of

peptides showed no difference in their relative abundance in the (+) and (-) ligand samples.

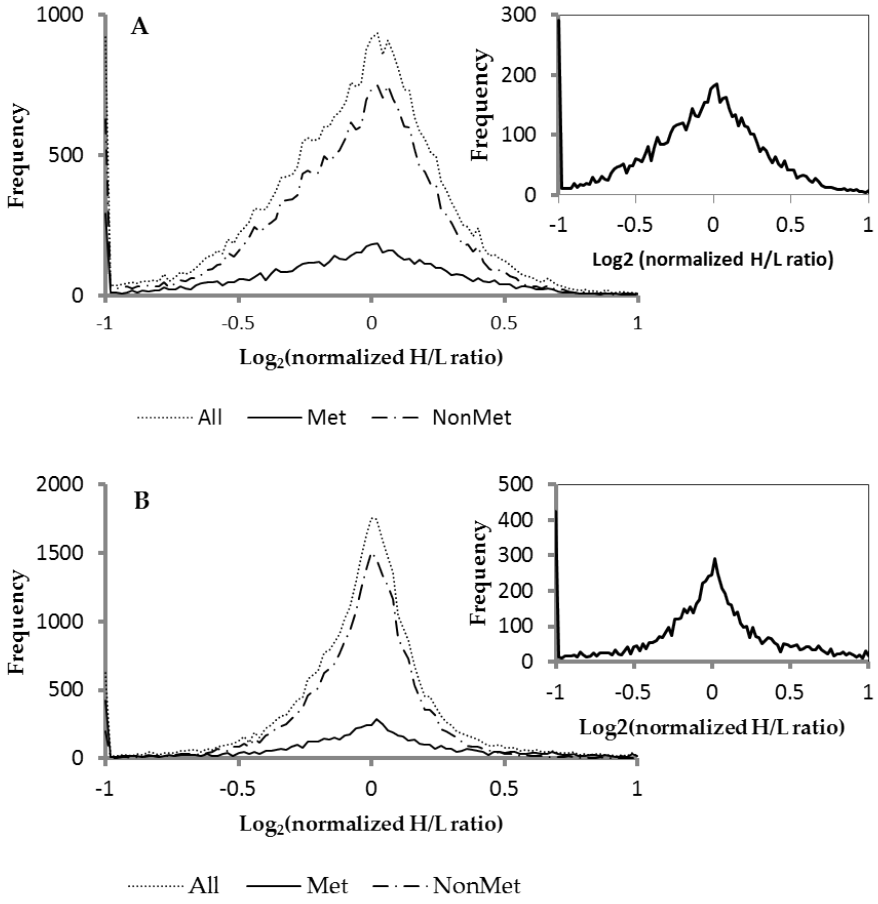

Figure 16: Distribution of the $\log_2$ of the normalized H/L ratios in (A) solution-based experiment 1B and in (B) solution-based experiment 2. Dotted lines represents distribution of all peptide sequences, dash-and-dotted line represents distribution of peptides that do not contain methionine in their primary sequences and solid line represents distribution of methionine containing peptides. Inset are zoom-in image of the methionine containing peptide distributions

A global analysis of the raw H/L ratios' distribution reveals approximately 98% of the SILAC-SPROX ratios lies within 0 to 10. These distributions show that the standard deviations (σ) of all peptides with H/L ratios from 0-10 lies within 0.32 -0.36.

Therefore, a significant H/L difference is defined as outside of [1 ± 2 σ] (equals to confidence intervals of 95%). That is, if one divides the altered H/L ratio at a particular denaturant concentration by the average H/L ratio across multiple denaturant concentrations for a given peptide, the result will be 1.7 fold or higher. This 1.7 fold difference is herein used as a threshold to filter for significant difference in the aforementioned data analysis routine.

In total, 302 proteins in the yeast cell lysate were effectively assayed for binding to the ATP analogue utilized in this study, and 180 peptide hits from 112 different proteins were identified. Increasing proteome coverage resulting from switching between QTOF and Orbitrap platforms results in significant increase of known ATP-binding proteins assayed in these solution-based experiments. Known ATP-binding proteins are defined as protein annotated with the GO-term ATP-binding (GO:0005524) from the Yeast Genome Database (SGD). A search for the GO-term "ATP-binding" in the Yeast Genome Database results in 666 genes, accounting for 13% of the yeast verified open reading frames (ORFs). A total of 82 of these 666 genes were effectively assayed in these solution-based experiments, and 33 of these known ATP binding proteins were detected as hits. The large majority of these hits were identified in experiment 2 where the higher concentration of the ATP ligand made the assay more sensitive.

4.3.2.2 Representative SILAC-SPROX data from Phosphoglycerate mutase

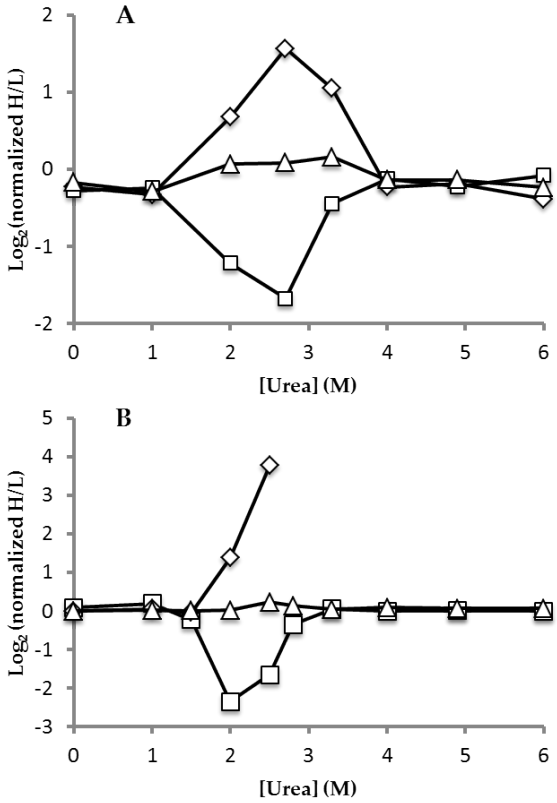

Figure 17: Representative SILAC-SPROX data from phosphoglycerate mutase (PGM-1) in (A) Solution-based experiment 1B, and (B) Solution-based experiment 2. Diamond shape represents data points from the un-oxidized methionine containing peptide (TVMIAAHGNSLRGLVK); square shape represents data points from the oxidized methionine containing peptide (TVM(ox)IAAHGNSLRGLVK) and triangle shape represents data points from a selected non-methionine containing peptide (LSRAIQTANIALEK)

Shown in Figure 17 is SILAC-SPROX data obtained on Phosphoglycerate mutase 1 (PGM-1). PGM-1 homolog was identified as a tentative ATP-binding protein in previous study using the energetics based pulse proteolysis approach in *E. coli* (See Introduction- ATP binding discovery). PGM-1 was not annotated with the GO-term

ATP-binding and was not previously known to binding ATP. In the previous ATP-binding in *E. coli*, the PGM-1 homolog was cloned, purified and validated to bind ATP.

In this ATP-binding study using the solution-based SILAC-SPROX described here, PGM-1 was identified as hits in both experiments. The data from both oxidized and un-oxidized Methionine containing peptides in solution-based experiment 1B resembles the structure of SILAC-SPROX curves predicted in Figure 14 (Data Analysis). Note that the SILAC-SPROX curve from the un-oxidized (TVMIAAHGNSLRGLVK) is a mirror image of the oxidized (TVM(ox)IAAHGNSLRGLVK) methionine containing peptide about the x axis. The straight line represents SILAC-SPROX data from a typical non-methionine containing peptide which is not affected by the oxidation reaction and is expected to center around 0. PGM-1 is also identified as one of the 3 hits detected in solution based experiment 1A on the QTOF platform and showed the same behavior as that in Figure 14 (data now shown).

This data is also in agreement with previous data obtained on the *E. coli* homolog, in which the C_m value of PGM increased from 1.4 M to 1.8 M Urea upon ATP-binding. Reconstruction of SPROX curves from both of these solution-based SILAC-SPROX data using a procedure reported elsewhere[63] indicates a $C_{1/2}$ shift of 1 M Urea from 1.8 to 2.8 M Urea (see Table 11 and Appendix B). Endogenous PGM-1 is a dimer in solution, and thus the thermodynamic stability is dependent on protein concentration. Thus, small discrepancy in relative apparent C_m shifts is to be expected due to changes

in protein concentrations. Interestingly, the results reported here for yeast PGM indicate that the protein folding and ATP-binding properties of this protein are very similar to that of *E. coli* PGM. This is despite that fact that the primary amino acid sequences of the two proteins are only 54% conserved.

4.3.2.3 Representative SILAC-SPROX data from GAPDH

As noted above in the Introduction to this Chapter, GAPDH was previously found to be destabilized in the presence of ATP [21a, b]. In theory, if a ligand binds to the native state of a protein, the protein is expected to be stabilized in the presence of the ligand [6a, 7-8]. Park and coworkers performed a number of biophysical studies on GAPDH binding to ATP and concluded that ATP binds and stabilizes a partially unfolded intermediate of GAPDH rather than native GAPDH resulting in an apparent destabilization[64]. For example, using tryptophan fluorescence emission spectroscopy, GAPDH showed a shift in apparent C_m by 0.2M Urea from 2.0 to 2.2 and decrease in m value from 3.2 to 1.7 (by half). Monitoring the tryptophan fluorescence intensity on the other hand showed two transitions in the unfolding behavior of GAPDH in the presence of ATP; with one transition at about 1.6 M and the other about 2.4 M Urea. Using ANS, a compound that tends to bind to nonpolar surface exposed in partially unfolded proteins and fluorescence, the Park group found the presence of an intermediate at approximately 2 M Urea in the presence but not in the absence of 1 mM ATP. In addition using size exclusion gel filtration, Park and coworkers found that GAPDH was

predominantly tetrameric at 2M Urea in the absence of ATP and mostly dimeric at 2M Urea in the presence of ATP. Park and coworkers also showed that ATP does not bind to native form of *E. coli* GAPDH but a seemingly dimeric partially unfolded intermediate with an estimated K_d of 150 μM. The effect of ATP binding to GAPDH is increasing both the folding and unfolding rate of the protein. They also went on and determined the effect of other nucleotides binding to GAPDH and saw the same results with ADP and AMP.

More recently, a book discussing the biological properties of GAPDH in detail was published and also indicated the binding of ATP to the dimer form of yeast GAPDH, which results in promoting the tetramer to dimer dissociation reactions[65]. This study also showed that binding of ATP to GAPDH at low temperature (i.e. 0^oC) results in almost complete loss of activity after 5 hours. The estimated K_d from this study in yeast (450 μM) was slightly higher than that in the *E. coli* study (150 μM). Also according to yeast GAPDH study, the ATP binding site lies in the N term region with sequence "GXGXXG" for a number of organisms. Yeast GAPDH exists in 3 isoforms with very high sequence identity (more than 90%) and share 70% conserved sequence with their *E. coli* counterpart. An alignment of the *E. coli* GAPDH sequence with the Yeast GAPDH-3 sequence gives rise to a highly conserved region on the N-term "INGFGRIGR", which resembles the typical ATP-binding motif.

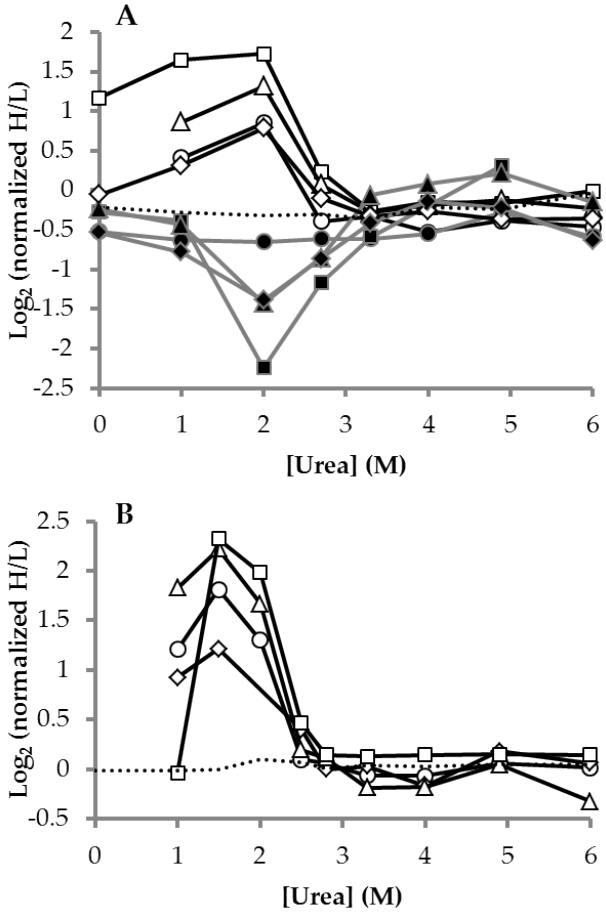

Figure 18: SILAC-SPROX data for multiple peptides from GAPDH in Solution-based experiment 1B (A) and Solution-based experiment 2 (B). Circles are peptides with sequence NVEVVALNDPFISNDYSAYMFK; triangles are VINDAFGI-EEGLMTTVHSLTATQK; diamonds are LTGMAFRVPTVDVSVVDLTVK; and squares are (K)VVITAPSSTAPMFVMGVNEEK. Closed symbols represent un-oxidized and open symbols represent oxidized methionine containing peptides. Dotted line represents SILAC-SPROX data from a non-methionine containing peptide (VLPELQGK).

The SILAC-SPROX results reported here on yeast GAPDH are consistent with the results previously reported by Parks and coworker on *E. coli* GAPDH (see Figure 18 below). Interestingly, multiple peptides from GAPDH in both solution-based SILAC-

SPROX experiments indicate the same change in thermodynamic stability of the protein upon binding to ATP, namely apparent destabilization.

These results altogether suggest the following (i) ATP (AMP or ADP) binds to a partially unfolded dimeric form of GAPDH (ii) this binding results in dissociation of tetrameric GAPDH (iii) The folding and ATP-binding properties of the yeast GAPDH isoforms are similar to each other and to their *E. coli* homolog and (iv) the ATP-binding site seem to be localized in the N-term region but also have remote effect on other domains of the protein. This result is very intriguing because it demonstrates the strength of these energetics based approach to identify not only binding that results in association but also dissociation of a protein complex, which would not be possible, for example, for ligand affinity and chemoproteomic target identifications (see Introduction).

4.3.3 False Positive/Negative Rate

4.3.3.1 False Positive Rate

In these solution-based ATP binding studies using SILAC-SPROX, it is possible to ascertain the false positive rate by examining the data obtained on non-methionine containing peptides (i.e. peptides that contains no Methionine in their primary sequence), which should not be affected by the oxidation reaction performed in SPROX analysis. Thus, the H/L ratios of these non-methionine containing peptides should be close to 1 and should not change regardless of which denaturant containing buffer they

come from. However, it is important to note that due to small differences in protein expression level between the Light and Heavy lysates, the non-Methionine containing peptide H/L ratios can fluctuate lower or higher than 1 depending on particular proteins. In theory, if the non-Methionine-containing data set is subject to the same data analysis routine adopted for the Methionine containing peptides, there should be no emerging peptide hits. Any protein hits found in such an analysis will be an indication of the false discovery rate of the data analysis. When the 1898 non-methionine containing peptides from solution-based experiment 1B were subjected to the same data analysis routine described in experimental procedures; a total of 24 non-methionine containing peptides satisfied the requirement to be a potential hit. In solution-based experiment 2, 1739 non-methionine containing peptides are assayed (i.e. have H/L ratio ≥0, appear in 8 denaturant concentrations or more), 24 peptides were also identified as hits. These analyses suggest a false-positive rate of protein target identification of 1.3% of the number of peptides assayed.

It is important to note that misidentification of peptides in the LC-MS/MS analysis step of SILAC-SPROX experiments can lead to the presence of false positives in the SILAC-SPROX experiment. In this regard, the false discovery rate of protein-ligand targets in the SILAC-SPROX experiment should not be lower than the FDR of the database search result, which was set to ≤1% because there is a 1% chance that a particular sequence is a mis-identified one.

4.3.3.2 False Negative Rate

The false negative rate associated with SILAC-SPROX experiments is more difficult to determine and it is likely to depend on the system under study. There are several caveats to the use of SILAC-SPROX in protein-ligand binding experiments. One caveat is that the ligand binding event must shift the SPROX transition midpoint by a measurable amount. The magnitude of a SPROX transition midpoint shift is related to the free ligand concentration, binding affinity, and protein folding m-value. The hydrogen peroxide concentration and reaction time can also impact the magnitude of the shift. The use of more aggressive oxidation reaction conditions (e.g., longer reaction times and/or higher concentrations of hydrogen peroxide) can make the assay more sensitive to the detection of weaker binding interactions (see discussion in Chapter 2 and Figure 2). It is also important that the protein (or protein domain) contains a buried methionine residue. Furthermore, oxidation of the methionine residue can abrogate ligand binding.

4.4 Conclusions

In conclusions, the SILAC-SPROX protocol reported here has been applied to the study of ATP-binding proteins in yeast. The protocol has eliminated several drawbacks associated with the current iTRAQ-SPROX such as (i) the high experimental error associated with separate sample handling prior to multiplexing by iTRAQ reagent (ii) the reliance on MS2 quantitation. First, the SPROX generated protein samples from

SILAC-based protocol reported here involved upfront labeling of the (-) and (+) ligand samples at protein level allowing combination of the two samples, eliminating discrepancy arising from sample handling procedure prior to LC MS/MS analysis. Second, the use of SILAC-based quantitation allows confirmation of the peptide (protein) hits by consulting the raw MS1 spectra. As a result, the SILAC-SPROX protocol described here has an improved false discovery rate of the SPROX analysis from 2-4 % to 1.3%. The protocol has successfully assayed 302 proteins and found 112 potential ATP-binding targets; of which 33 are previously annotated with GO-term ATP-binding in the Yeast Genome Database. The preliminary results are consistent with ATP-binding studies in *E. coli*. The proteome coverage is also close to that reported in the E. coli study using the pulse proteolysis approach. The proteomic coverage is also similar to that observed in a previously reported iTRAQ-SPROX experiment in which 327 yeast proteins were effectively assayed for binding to Cyclosporine A[35a]. This is not surprising because the solution-based SILAC-SPROX protocol described here and the iTRAQ-SPROX protocol both rely on the identification and quantitation of the methionine containing peptides. However, this reliance on methionine-containing peptides in the solution-based SILAC-SPROX protocol can be eliminated using a gel-based SILAC-SPROX protocol as described below.

5. Development of a SILAC-SPROX-Cyanogen Bromide protocol and application to ATP binding discovery

5.1 Motivation

The SILAC-SPROX-Cyanogen Bromide protocol described in this chapter is an extension of the SILAC-SPROX protocol described in chapter 4. The SILAC-SPROX protocol described in Chapter 4 eliminated some of the drawbacks associated with the iTRAQ-SPROX protocol. For example, it allows the (+) and (-) ligand samples to be combined early in the protocol (i.e., immediately after SPROX analysis), rather than late in the protocol (i.e.., after the proteolytic digestion step). However, the SILAC-SPROX protocol described in Chapter 4 still requires the detection and quantification of Methionine containing peptides. As methionine peptides only account for 20% of identified peptides in a typical proteomics experiment, the proteome coverage in the SILAC-SPROX experiments is roughly 1/5 of that observed in conventional bottom-up proteomic analyses. In theory, enrichment of Methionine containing peptides can be used to improve the proteome coverage in SILAC-SPROX experiments. However, it is time consuming and costly to perform 8-10 different Methionine enrichments before subjecting SILAC-SPROX samples to LC MS/MS Analysis.

Described in this chapter is a SILAC-SPROX-Cyanogen Bromide (SILAC-SPROX-CnBr) strategy that is designed to expand the protein coverage in proteome-wide SPROX experiments. The SILAC-SPROX-CnBr strategy enables any peptide (i.e., methionine-containing or not) that is identified and quantified in a bottom-up shotgun

proteomics experiment to report on the stability of the protein to which it maps. As part of the work described here an ATP-binding study is performed using the same two conditions employed in the solution-based experiments described in Chapter 4. The results obtained from this study including the increased proteome coverage observed and the ATP binding properties of selected yeast proteins will be discussed.

5.2 Experimental Procedures

5.2.1 SILAC-SPROX Analyses

Table 9: Experimental parameters utilized in ATP-binding gel-based experiment 1 and 2

Experiment Name	*MS platform*	*[Ligand]*	*Equilibration Time (min)*	*Oxidation Time (min)*	*Number of Denaturant concentration*
Gel based Exp1	Orbitrap	1mM	30 and 30	6	8
Gel based Exp2	QTOF	7mM	60 and 60	24	10

Two SILAC-SPROX-CnBr experiments were performed. Summarized in Table 9 are key experimental conditions employed for these experiments (i.e. the Gel-based experiment 1 and 2). Note that Gel-based Experiment 1 employed the same ATP-binding and SILAC-SPROX Analysis as the Solution-based Experiment 1 A/B described in Chapter 4. Gel-based Experiment 2 also employed the same ATP-binding and SILAC-SPROX Analysis as the Solution-based Experiment 2 in Chapter 4.

5.2.2 Cyanogen Bromide Digestion

In Gel-based experiment 1, the combined (-) and (+) ligand protein pellets were re-dissolved in 70% (v/v) aqueous solution of formic acid (TCI America, Portland, OR) containing 1 mg of crystalline CnBr (Sigma Aldrich) and incubated with frequent shaking at room temperature for 4 hours. This was equivalent to an estimated CnBr: Methionine ratio of ~ 67:1 (mole/mole). The unreacted CnBr was evaporated by heating samples with open caps at 50^{o}C for 5 min in chemical fume hood. The protein samples were neutralized by 200 μL of 1.7 M of 4-ethylmorpholine (NEM) (Sigma Aldrich) and diluted with 1.5 mL of diH_2O and 200 μL of 100% (w/v) TCA was added to each sample. The samples were mixed and incubated at 4^{o}C overnight. Precipitated proteins were pelleted by centrifuging at 13,000 g for 30 min, and washed according to similar procedures described in the SILAC-SPROX protocol in Chapter 4.

The CnBr digestion in Gel-based Experiment 2 was performed in a similar manner as that described above for Gel-based Experiment 1 with the following exceptions (i) a 5 M solution of CnBr in Acetonitrile (Sigma Aldrich) was used instead of crystalline CnBr, and (ii) the final estimated CnBr: Methionine Ratio was ~ 138:1 instead of 67:1.

Important precautions! Cyanogen Bromide is a severely irritating chemical that has high acute toxicity. The lethal concentration of CnBr by inhalation for human is 92 ppm (398 mg/m^3 for 10 min) [66]. The amount of CnBr utilized in this work is

approximately 20-30 mg per SPROX experiment, therefore can be safely handled under a functional chemical fume hood. When weighing crystalline CnBr, be sure to use a closed container (e.g. a closed cap eppendorf tube); and never handle CnBr outside of the chemical fume hood. CnBr containing waste can be collected into a 1 L glass container with cap prefilled with 300 mL of household bleach and placed into a secondary container (e.g. a bucket). Make careful precautions when handling the waste; be sure to seal the cap, label the content of the waste container and have it picked up by authorized waste managers. Protective equipment such as lab coats, goggles and impermeable gloves should be worn at all time while working with CnBr.

5.2.3 1-D SDS PAGE and In-gel Digestion

A total of 8 samples for SDS-PAGE analysis were generated in Gel-based experiment 1 after combining (-) and (+) ligand samples (i.e. Light and Heavy lysates). Each of the 8 TCA precipitated protein pellets, which contained both the (-) and (+) ligand samples (i.e. Light and Heavy lysates), were re-dissolved in 40 μL of freshly made 8M Urea and 20 μL of 6X Laemmli sample buffer containing 375 mM Tris•HCl pH 6.8, 6% SDS, 50% Glycerol and 0.045% Bromophenol (Boston Bioproduct). A 3 μL aliquot of β Mercaptoethanol (BME) was added to each sample before heating at 95^{o}C for 5 min to reduce the disulfide bonds. A 20 μL aliquot of each protein sample was loaded on a mini polyacrylamide gel 10 x 8.5 cm (NUsep). The gel was fixed with fixing solution containing 25% isopropanol, 10% acetic acid in 65% distilled H_2O for 20 min. The fixed

gel was stained overnight with staining solution containing 0.01% R-250 in 10% Acetic acid (Bio-Rad), and de-stained with 10% acetic acid repeatedly until the gel image was clear. A portion of the gel corresponding to a molecular weight range of 20 – 30 KDa was excised resulting in 8 different gel bands, which were subject to a standard in-gel digestion protocol[67]. Gel-based experiment 2 was performed in a similar manner with the following exceptions (i) there were 10 samples after combining (-) and (+) ligand samples instead of 8; (ii) each samples was re-dissolved in 40 μL of freshly made 10M Urea instead of 8M Urea; (iii) 15 μL of each protein sample was loaded onto a midsize polyacrylamide gel (Bio-Rad Criterion); and (iv) the gel was cut into 14 different molecular weight (MW) fractions that correspond to the following MW ranges: 0-6, 6-12, 12-15, 15-18, 18-19, 19-20, 20-24, 24-30, 30-37, 37-50, 50-75, 75-100, 100-250 and larger than 250 KDa. The resulting 140 gel pieces were each subject to a standard in gel digestion as described elsewhere [67]. The Gel-cutting strategy for Gel-based experiments is depicted in Figure 19.

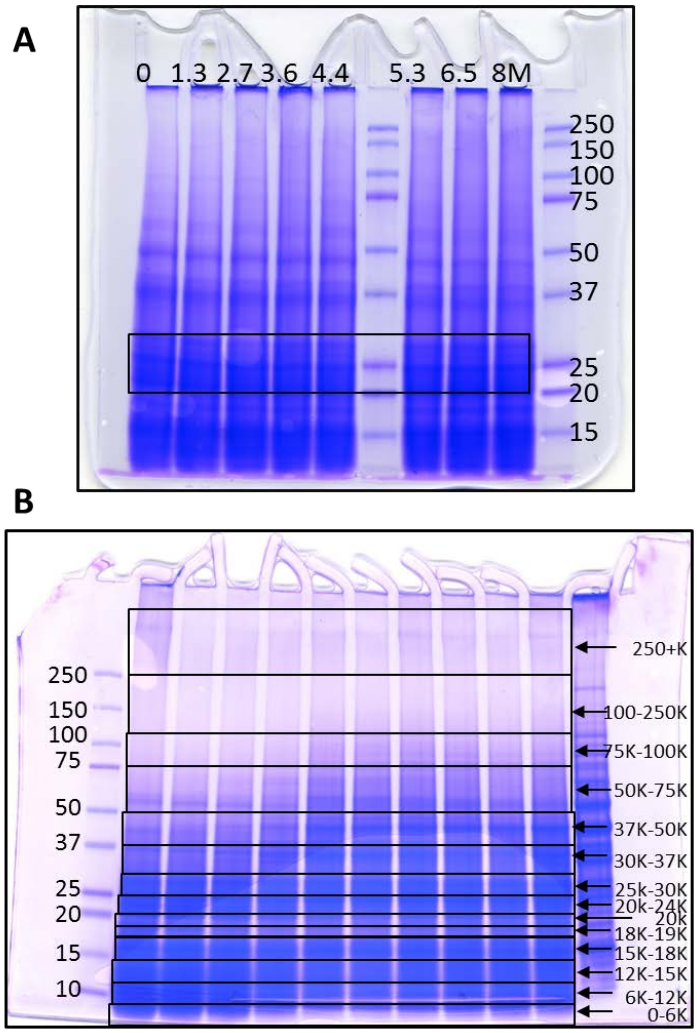

Figure 19: Gel-cutting strategy for (A) Gel-based experiment 1 and (B) Gel-based experiment 2. Black boxes represent relative sizes of the gel bands. Arrows indicates the estimated molecular weight ranges for each gel band.

5.2.4 LC MS/MS Analyses

The extracted peptide mixture from each in-gel digestion was evaporated down in a speed-van to approximately 100 µL, diluted with 500-1000 µL of buffer A and desalted using C18 resin. Peptides were eluted in 150 µL of 70% Acetonitrile in 0.1%

TFA before being speed-van down to dryness and re-dissolved in 20-50 μL of Buffer A (0.1% formic acid). Approximately 2-4 μL of each sample was combined with 19-20 μL of buffer A and subjected to LC-MS/MS analyses. The 8 SILAC-SPROX-CnBr samples from Gel-based Experiment 1 were subject to LC MS/MS Analysis on the Orbitrap platform in the same manner as the Solution-based experiment 1B and 2 described in Chapter 4.

The 10 SILAC-SPROX-CnBr samples in the Gel-based experiment 2 were subjected to an LC-MC/MS analysis using an Agilent 6520 Q-TOF mass spectrometer equipped with a Chip Cube interface. The HPLC column was a short chip with a 40 nL enrichment column and a 73 x 43 mm analytical column packed with Zorbax 80SB-C18 5μm material. Peptides were eluted using a linear gradient: 3-5% buffer B over 2 min, 5-15% over 2 min, 15-50% over 18 min 60-90% over 3 min, 100% over 2 min and 5% over 3 min. The flow rate was set to be 0.4 μL/min. The capillary voltage ranged from 1800 to 1850 V. The flow rate of the drying gas was set at 6 L/min at 350°C. The skimmer and fragmentor were set at 65 and 175 V, respectively. The collision energy was as determined by the equation 3.5 V/100 Da with a -4.8 V offset. The inclusion window width for precursor ions was 4 m/z. The scan rate was three scans per second in the mass spectra and two scans per second in the product ion spectra. In every cycle, four precursors were selected for fragmentation. All spectra were collected in profile mode.

5.3 Results and Discussion

5.3.1 General Strategy

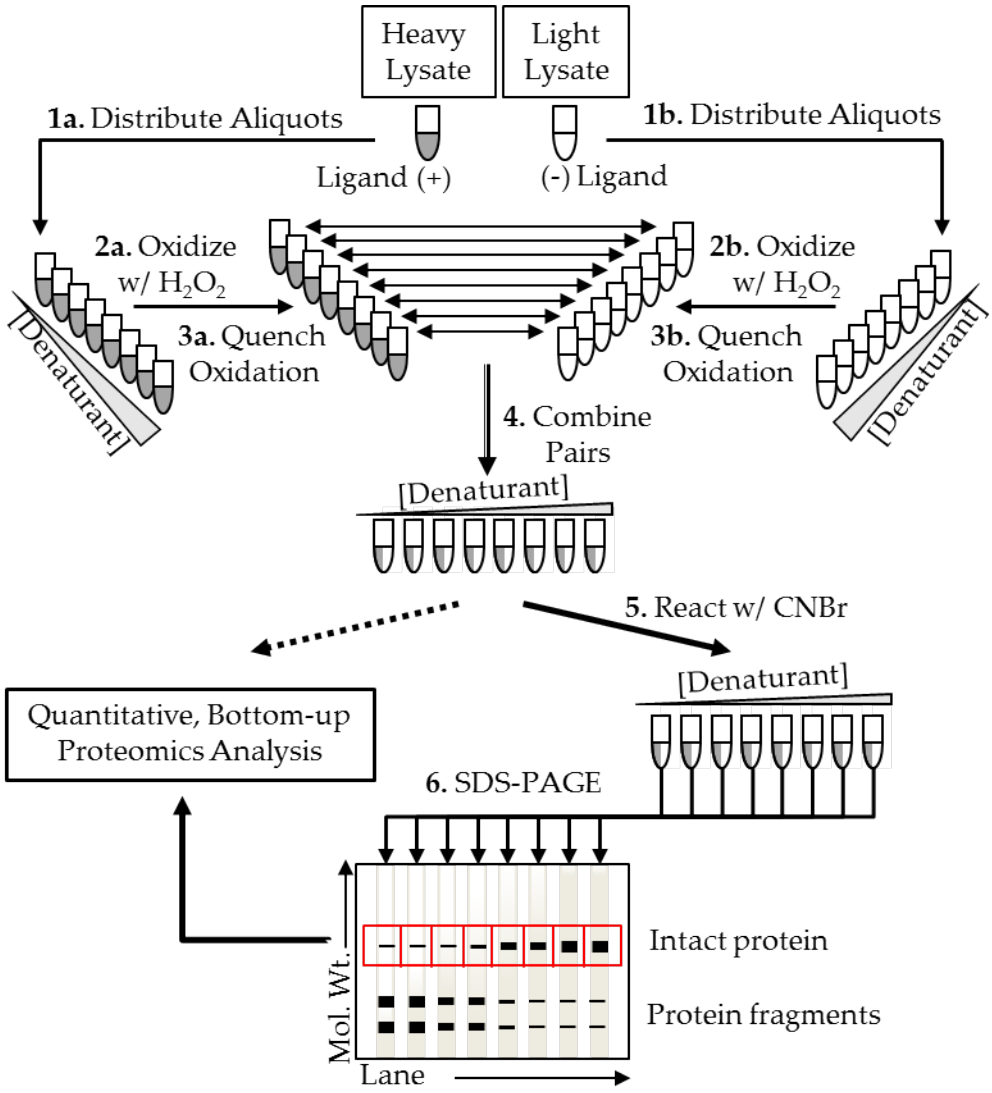

Figure 20: The SILAC-SPROX-Cyanogen Bromide Protocol

The SILAC-SPROX-CnBr protocol developed in this chapter is outlined in Figure 20. The first 4 steps of the protocol are identical to the solution-based SILAC-SPROX protocol outlined in Chapter 4. The CnBr reaction in step 5 of the SILAC-SPROX-CnBr

protocol is used to cleave the polypeptide chain of proteins after un-oxidized methionine residues (i.e., methionine residues that were not oxidized in the SPROX experiment). Methionine residues that were oxidized in the SPROX experiment are protected from CnBr cleavage. As proteins are unfolded in the presence of the increasing concentrations of chemical denaturant used in the SPROX experiment, the "buried" methionine residues in a protein's three-dimensional structure are exposed to solvent, get oxidized, and become protected from cyanogen bromide cleavage. Thus, as depicted in Figure 20, full-length and fully oxidized proteins (i.e. whose all Methionine residues in primary sequence are quantitatively oxidized) will appear predominantly in SPROX samples with higher concentrations of denaturant, and the corresponding CnBr fragments of proteins will appear predominantly in SPROX samples with lower concentrations denaturant. Therefore, a fractionation step by 1-D SDS-PAGE is used to separate full-length proteins from their corresponding CnBr fragments. Subsequently, rows of gel-bands corresponding to specific molecular weight ranges are excised, and subjected to conventional bottom up proteomics to quantify, in each gel-band, the relative amount of Full-length protein (or corresponding CnBr fragments) from the (-) and (+) ligand samples.

The CnBr digestion and subsequent gel-based separation of the full-length proteins from their respective CnBr fragments is important in the gel-based approach because it enables every peptide that is successfully identified and quantified in the

bottom-up proteomics readout (not just the methionine-containing peptides) to report on the ligand binding properties of the protein from which it was derived. However, the SILAC-SPROX behavior of full-length proteins and their respective CnBr fragments will be opposite of each other. The amount of a full-length protein will increase as a function of denaturant concentration; whereas, the amount of CnBr fragments of a protein will decrease as a function of denaturant (see Figure 20). Therefore, it is important that the gel be excised in molecular weight ranges that effectively separate intact proteins from their CnBr fragments.

Ultimately the H/L (or L/H) ratios of the identified peptides are evaluated as a function of the denaturant concentration at which the protein oxidation reaction was performed. Unique to the gel-based experiment is that the denaturant dependence of the H/L ratios obtained from non-methionine-containing peptides can also be used to identify protein hits. As with the methionine-containing peptides from protein hits in the solution-based experiment, the non-methionine-containing peptides from protein hits in the gel-based experiment will have high (or low) H/L ratios at intermediate denaturant concentrations (see Figure 21).

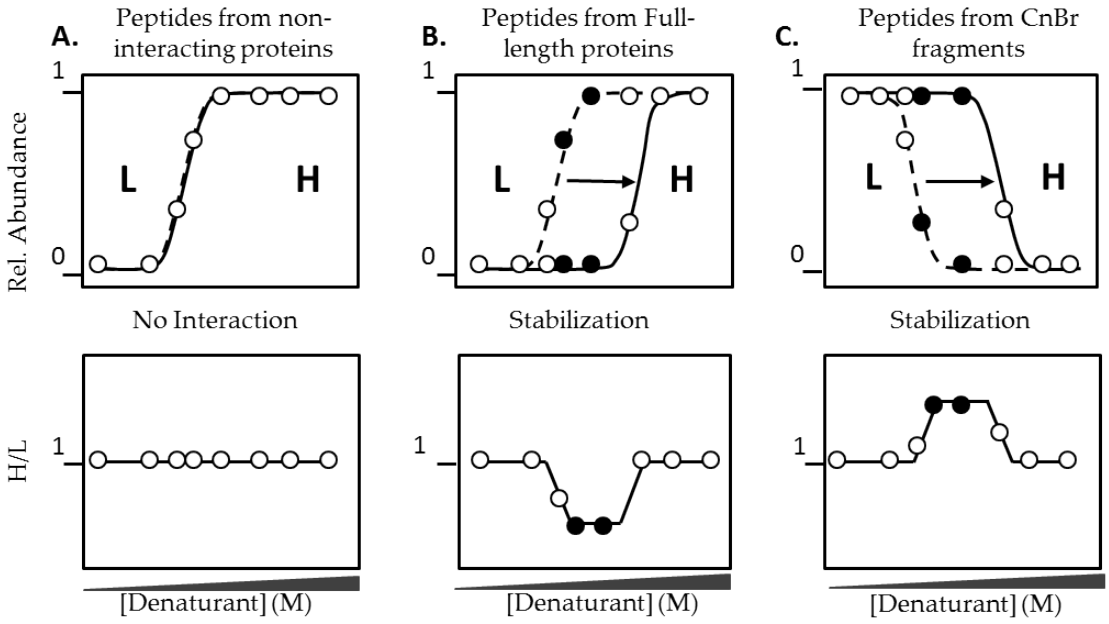

Figure 21: Expected SILAC-SPROX-CnBr results from the gel-based experiments; (A) is an oxidized methionine peptide of a protein that has no interaction with the ligand; (B) is any peptides from the full-length protein that is stabilized by binding to the ligand; and (C) is any peptides from the corresponding CnBr fragments of stabilized proteins. Open circles represent data points (denaturant concentrations) that have no change in H/L ratio, closed circles represent data points that have significant H/L ratio difference.

More specifically, any peptides that originate from the full-length proteins (regardless of whether or not they contain a methionine residue) will have similar SILAC-SPROX behavior as the oxidized methionine residues that the protein bears. Conversely, any peptides which originate from the CnBr fragments will have similar SILAC-SPROX behavior as the un-oxidized methionine residue that was cleaved by CnBr. Hence the so-called "full-length" originated peptides will have the reverse SILAC-SPROX curve as oppose to the "CnBr-fragment" originated peptides.

5.3.2 Proteome Coverage

Table 10: Proteome coverage from all ATP-binding experiments including solution-based 1A/B and 2; gel-based 1 and 2.

Experiment	*Total Peptides (Proteins) Assayed for Binding*	*Hit Peptides (Proteins)*	*Known ATP Binding Proteins Assayed (Hits)*
Solution-Based Exp. 1A	93 (38)	5 (3)	6 (0)
Solution-Based Exp. 1B	526 (209)	55 (27)	56 (4)
Gel-Based Exp. 1	1346 (354)	45 (12)	73 (1)
Solution-Based Exp. 2	353 (216)	138 (99)	61 (29)
Gel-Based Exp. 2	431 (171)	148 (46)	27 (8)
Total	2035 (526)	325 (140)	109 (37)

Presented in Table 10 is the proteomics coverage from all SILAC-SPROX-CnBr ATP-binding experiments performed in this work. Results from the 2 solution-based SILAC-SPROX experiments described in Chapter 4 are also included in the table for comparison. In Gel-based experiment 1, only bands from the 20 to 30 KDa molecular weight range of the gel were excised, digested and the resulting peptides subjected to an LC MS/MS Analysis. Even just considering the peptide and protein identifications from this 20-30 KDa molecular weight range, 2.5 times as many peptides and 1.7 times as many proteins are assayed compared to the Solution-based experiment 1B (i.e. 1346 vs. 526 and 354 vs. 209). This improvement is due to (i) removal of the requirement that a

methionine containing peptide must be identified and quantified in order for the proteins to be assayed and (ii) use of the 1-D gel fractionation strategy to reduce the complexity of the lysate samples prior to LC MS/MS Analysis.

The 140 SILAC-SPROX-CnBr samples generated in Gel-based experiment 2 were analyzed on an Agilent Q-TOF mass spectrometry platform. Due to the relatively low complexity of the in-gel digested samples, each SILAC-SPROX-CnBr generated samples were subject to a short (i.e. 30 min) LC MS/MS analysis using a short (75 μm x 43mm) HPLC column. A direct comparison can be drawn between the solution-based experiment 1A and the gel-based experiment 2 in terms of peptide and protein coverage. The results suggest that SILAC-SPROX-CnBr increased the proteome coverage by 5 fold compared to the SILAC-SPROX protocol (i.e. from 38 proteins to 171 proteins).

The analytical capabilities (e.g., speed and sensitivity) of the mass spectrometry instrument used in the SILAC-SPROX experiment clearly impact the proteome coverage. The total number peptides (and proteins) assayed using the Orbitrap instrument in Solution-Based Experiment 1B, 526 peptides (and 209 proteins), was about 5-fold greater than that obtained using the Q-TOF instrument in solution-based experiment 1A, 93 peptide (and 38 proteins). A comparison of the protein coverage in Gel-Based Experiment 1 (1346 peptides and 354 proteins) to the protein coverage obtained in the 20 to 30 KDa molecular weight range of the gel in Gel-Based Experiment 2 (237 peptides and 102 proteins) also revealed a similar 3 to 5-fold increase in peptide and protein

coverage when the Orbitrap instrument was used in the gel-based protocol. The proteome coverage obtained in the large-scale gel-based experiment using the Q-TOF (i.e., 431 peptides and 171 proteins in Gel-Based Experiment 2) was also ~5-fold larger than that of the solution-phase experiment using the Q-TOF (i.e., 93 peptides and 38 proteins in Solution-Phase Experiment 1A). If a similar 5-fold improvement were realized using the Orbitrap instrument, a large-scale gel-based experiment using the Orbitrap would be expected to assay about 1800 to 2500 peptides and about 1000 proteins, which is five times the number of peptide and proteins assayed in Solution-Based experiments 1B and 2. This is approximately 3-5-fold more peptide and protein coverage than that previously reported in SPROX experiments using an isobaric mass tagging strategy[35b].

One important feature of the gel-based SILAC-SPROX protocol is that it increases the number of peptide probes per protein. In the two gel-based SILAC-SPROX experiments in this work there were 3.2 and 3.8 peptide probes per protein, whereas in the solution-based experiments described in Chapter 4, the number of peptide probes per protein was between 1.6 and 2.5. Moreover, the protein hits in the gel-based experiments were identified with more peptide probes. For example, the protein hits in the two gel-based experiments were identified with an average of 3.2 and 3.8 peptides each, whereas the protein hits in the solution-based experiments were only identified with an average of between 1.4 to 2 peptides each.

It is also noteworthy that the solution- and gel-based protocols are complementary in their peptide and protein coverage. Shown in Figure 22 is a comparison of the peptide and proteins assayed and ATP binding hits detected in all of the gel- and solution-based experiments described here.

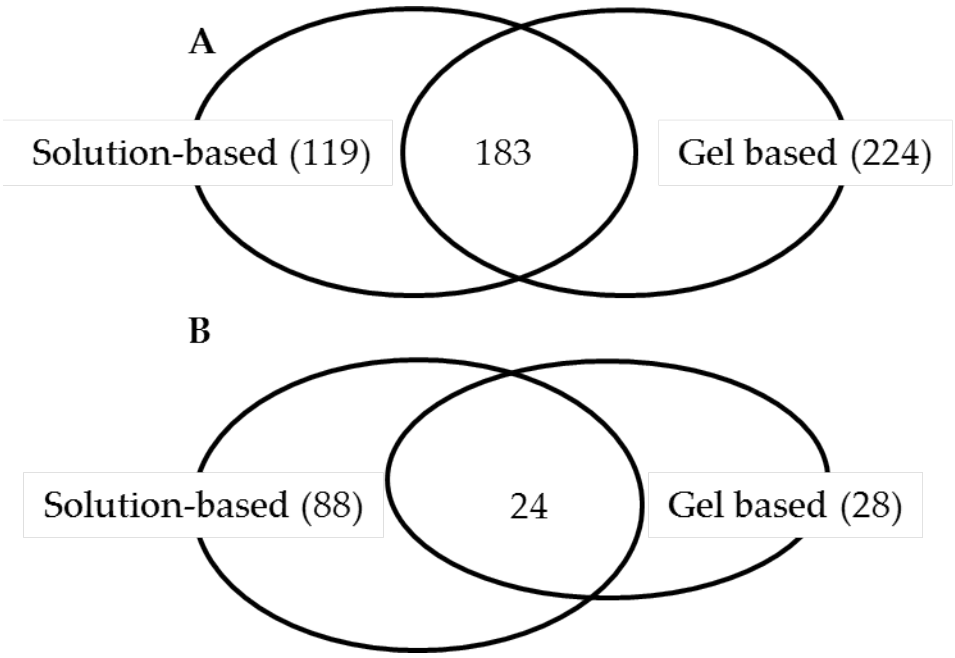

Figure 22: A comparison of the (A) proteome coverage (i.e. assayed proteins) from gel-based and solution-based experiments and (B) potential protein hits from gel-based and solution-based experiment

In total, 526 proteins in the yeast cell lysate were effectively assayed for binding to the ATP analogue covering a total of 109 proteins annotated with "ATP-binding" GO-term. A total of 325 peptide hits from 140 different proteins were identified (see Appendix B), and 37 of these known ATP binding proteins were identified as hits. These 37 proteins constituted 25% of the protein hits in this study. Many of the other protein hits in this study are also known to bind other nucleotides (e.g., NAD, GTP),

DNA, RNA and a number of ribosomal proteins were identified as hit proteins in this study (see Figure 23). It is also noteworthy that 6 of the newly discovered ATP binding proteins in this study are homologues to *E. coli* proteins that were recently reported to have ATP binding interactions based on results of a pulse proteolysis study[21b] (Data not shown).

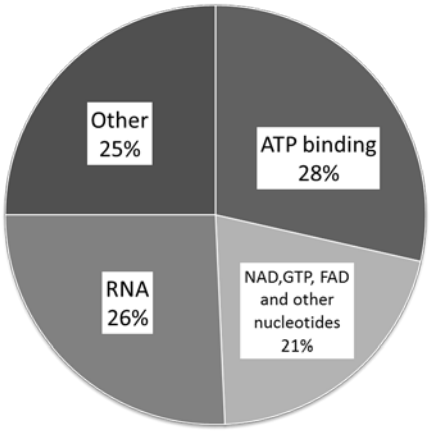

Figure 23: Distribution of known ligands for the hit proteins in this study

5.3.3 Representative SILAC-SPROX-CnBr data from PGM-1

In the SILAC-SPROX-CnBr experiment, the gel bands must be cut in molecular weight ranges such that the full-length protein and the cyanogen bromide fragments are not in the same set of gel-bands. This is demonstrated clearly in the case of PGM-1. PGM-1 is a small globular protein with 247 amino acids in its primary sequence (approximately 27 KDa in MW). It is known from previous solution-based SILAC-

SPROX experiments that the yeast PGM-1 binds to ATP and ATP-binding stabilizes the protein (See Chapter 4).

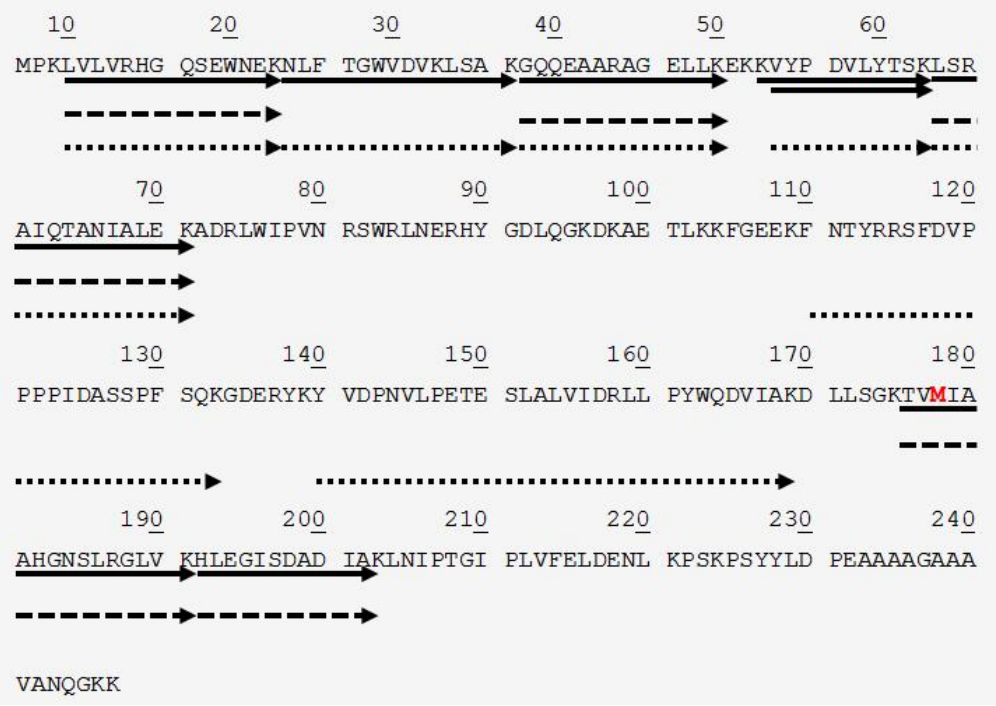

Figure 24: The sequence coverage of PGM-1 in the 2 gel-based experiments. Each arrow represents a peptide identified in the LC MS/MS. Solid arrows represent peptides identified the gel-based experiment 1. Dash arrows represent peptides identified in the gel-based experiment 2 and correspond to the full-length protein. Dotted arrows represent peptides identified in the gel-based experiment 2 and correspond to the CnBr-fragment.

PGM-1 has 2 Methionine residues, one is the initial methionine and the other is at position 180. CnBr cleavage at Met-180 results in two fragments; including one which is 21 KDa and one which is 6KDa. This makes PGM-1 an excellent model for this ATP-

binding Study. In the gel-based experiments, PGM-1 serves as a positive control for the ATP binding reaction, SPROX Analysis and also gel cutting strategy.

Presented in Figure 24 is the sequence coverage obtained for PGM-1 in both Gel-based experiment 1 and 2. In the Gel-based experiment 1, the gel was excised between 20 and approximately 30 KDa (See Figure 19, experimental procedures). This gel band thus contains both the fragment (21KDa) and the full-length protein (27KDa). As expected the sequence coverage of PGM-1 this first gel-based experiment showed: (i) peptides that are common between the 21 KDa fragment and the full-length protein; and (ii) peptides that are unique to the full-length protein (TVM(ox)IAAHGNSLRGLVK and HLEGISDADIAK). Peptides from the full-length protein are expected to have the opposite SILAC-SPROX curve as peptides from the 21 KDa CnBr fragment. Thus, peptides that are common to both the full length protein and the 21 KDa CnBr fragment (see Figure 24) will have SILAC-SPROX curves that are the sum of the two behaviors, resulting in H/L ratios close to 1 regardless of the denaturant concentration (see Figure 21 – Expected Results). Peptides that are unique to the full-length protein should have SILAC-SPROX curves resemble that shown in Figure 21, in which the H/L ratios displayed a significant dip in the transition region. This was indeed the case for all PGM-1 peptides obtained in the gel based experiment 1 (see SILAC-SPROX curves of a selected non-methionine containing peptide with sequence LSRAIQTANIALEK in Figure 25A).

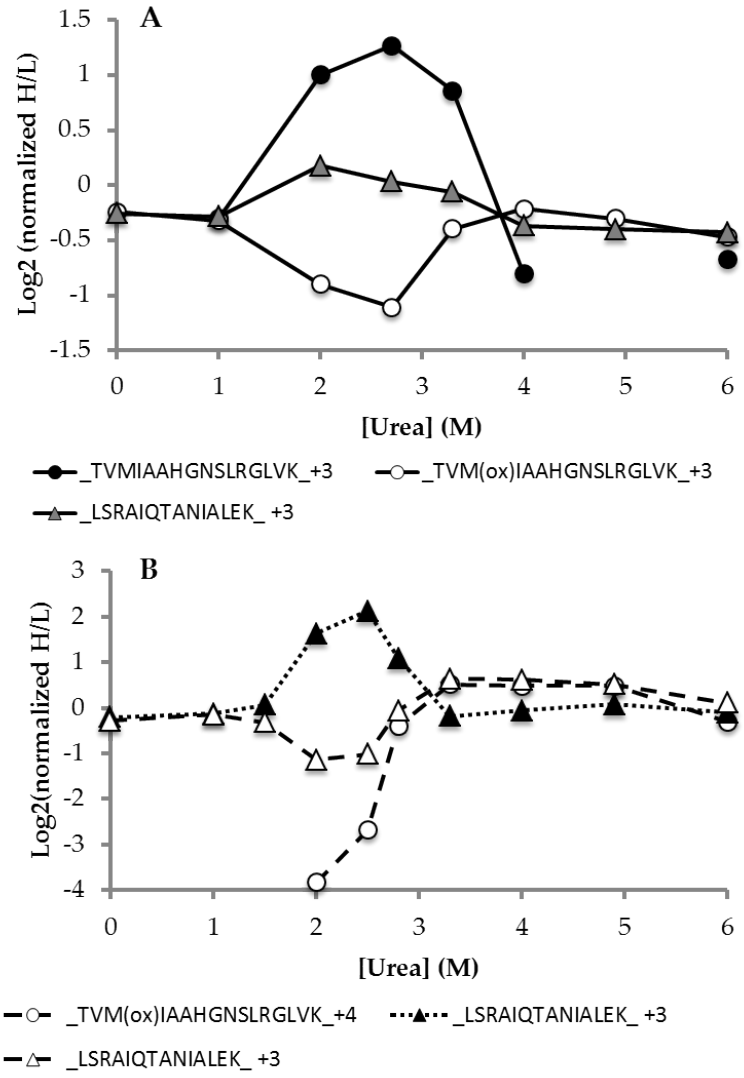

Figure 25: Representative SILAC-SPROX-CnBr data from PGM-1 in (A) gel-based experiment 1 and (B) gel-based experiment 2. Closed circles represents data points from peptide TVMIAAHGNSLRGLVK, open circles represents data from TVM(ox)IAAHGNSLRGLVK; closed triangles represents data points from a selected non-methionine containing peptide LSRAIQTANIALEK in the CnBr Fragment; open triangles are data points from LSRAIQTANIALEK in the Full-length protein. Solid line represents peptides from gel-based exp. 1; dotted line represents peptide from gel-based exp. 2 and the CnBr fragment; dashed line represents peptides from gel-based exp. 2 and the Full-length protein (see Figure 24).

Ultimately, because of the cutting strategy employed in gel-based experiment 1 (see Figure 19), the only peptide hits identified were (TVM(ox)IAAHGNSLRGLVK) and (TVMIAAHGNSLRGLVK) (See Figure 25A). Interestingly, the presence of TVMIAAHGNSLRGLVK in the peptide readout suggests incomplete cleavage of CnBr; this is probably the reason why the (HLEGISDADIAK), a peptide unique to the full-length protein did not show up as a hit in Gel-based experiment 1(Data not shown).

In Gel-based experiment 2, a more aggressive CnBr digestion was employed along with an improved cutting strategy (See Figure 19, experiment procedures). The cutting strategy in this region (also one of the most intensely coomassie-stained region) involved generating 2 sets of gel-bands; one set in the 20-24 KDa and the other in the 24-30 KDa MW range. This allowed the full-length 27KDa protein to be effectively separated from its corresponding CnBr Fragment (21 KDa).

The sequence coverage obtained in this second gel-based experiment (see Figure 24) showed no evidence of the un-oxidized methionine containing peptide, which is consistent with more complete cleavage by CnBr. In this case, all peptides from the full-length protein regardless of whether or not they are common or unique are expected to show a "dip" in the H/L ratios from the transition region (i.e. around 2 M Urea) (see Figure 21 - Expected Results). Conversely, all peptides from the CnBr fragment are expected to show a "peak" in the H/L ratios from the transition region (see Figure 21 –

Expected Results). This was indeed the case for all peptides obtained for PGM-1 in the gel-based experiment 2 (see Figure 25B and 26).

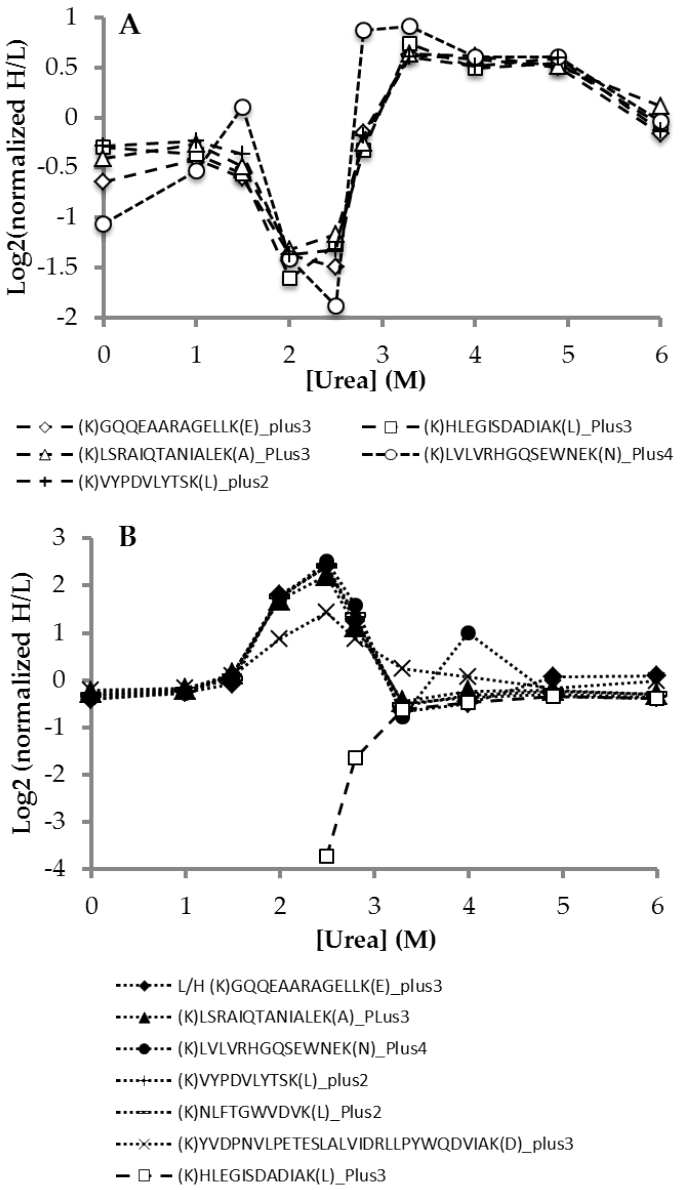

Figure 26: SILAC-SPROX-CnBr data of PGM-1 in gel-based experiment 2. (A) are all non-methionine containing peptides identified in Full-length protein and (B) are those identified in the 21 KDa CnBr Fragment. Dashed line and open symbols represents peptide originates from Full-length protein; dotted line and closed symbols represents peptides originates from 21 KDa CnBr Fragment.

Interestingly, the change in H/L ratios seems to be more dramatic for methionine containing peptide than that of non-methionine containing peptides (see Figure 25B). Also, the peptide that is unique to the 27KDa Full-length protein (HLEGISDADIAK) was also identified in the gel band with MW range of 20-24 KDa in the gel-based experiment 2 and showed opposite SILAC-SPROX curves from peptides that originate from the CnBr fragment (see Figure 26 below). These suggest that the cutting strategy may not separate all full-length protein from the CnBr-fragment. Nevertheless, the cutting used in this case did not affect the ability of the non-methionine containing peptides to report on the binding properties of the protein to ATP. This suggests that the cutting strategy in the SILAC-SPROX-CnBr experiments does not have to be perfect. An imperfect cutting strategy can be tolerated at least in some cases.

The cutting strategy clearly has an impact on the false negative rate in the Gel-based experiment 1. Although as many as 354 proteins are assayed (including 73 known ATP-binding proteins), only 12 proteins are identified as hits, including only one of the known ATP-binding proteins. It is also important to stress that the 20-30 KDa region is the most crowded region that contains a large fraction of the proteins/fragments on the gel; cutting this region into smaller gel pieces clearly seems to be beneficial. When performing such a 1-D fractionation strategy, it might be advantageous to choose alternative gels to better resolve the fragments in this 20-30 KDa region. Also impacting

the false negative rate in gel based experiment 1 is the use of less aggressive oxidation reaction conditions.

5.3.4 Representative SILAC-SPROX-CnBr data from Phosphoglycerate Kinase (3-PGK)

Phosphoglycerate kinase (3-PGK) is another ATP-binding protein identified in the gel-based experiments described here. As discussed in the histidine HDX study in Chapter 2, 3-PGK has 2 functional domains that fold independently from each other. The N-terminal domain has a $C_{1/2}$ determined from histidine HDX of 1.5 M GdmCl and the C-terminal domain has a $C_{1/2}$ value of 0.5 M GdmCl. These $C_{1/2}$ values can be roughly translated to 3 M and 1 M Urea, respectively. The ATP-binding region is presumably located in the C-terminal domain between residues 205-220 (PFLAIL**GG**A**K**VADKIQ)[68]. In Solution-based experiment 1 and 2, one methionine containing peptide was identified, (229)-VDSIIIGGGMAFTFK-(243). However this peptide was not effectively assayed in either case because it only appears in 5 SILAC-SPROX generated samples in Solution-based experiment 1 and 6 SILAC-SPROX generated samples in Solution-based experiment 2. As a result, 3-PGK was not identified as a hit in the solution-based experiments.

A visual inspection of the 3-D Structure of this protein (PDB: 1QPG) suggests this methionine (Met 238) is protected and in close proximity to the ATP-binding region. However the SILAC-SPROX results showed no change in the H/L ratios as a function of Urea concentrations in the range of 2.5 to 6 M indicating no interaction. Although it is

interesting to note that, the peptide was not identified in SILAC-SPROX samples with Urea concentration < 2.5 M. According to the data generated in the Histidine slow H/D exchange experiments (see table 3), the mid-point of this C-term domain is 0.5 M GdmCl, which is consistent with previous results in reference[69]. This $C_{1/2}$ value is approximately 1 M Urea and is outside of the 2.5 M to 6 M Urea range that was probed in the solution-based SILAC-SPROX experiment. It is noteworthy that 3-PGK is identified to be a ATP-binding hit in Gel-based experiment 2 in which non-methionine containing peptides can be used to report on the thermodynamic stability of proteins. In fact, all 14 peptides that were identified from 3-PGK in the 37-50 KDa MW range of the gel-band in Gel-based experiment 2 had a denaturant dependence to their H/L ratios, which consistently indicates that 3-PGK is an ATP binding protein (see Figure 27).

3-PGK has 3 methionine residues; including one that is located on the N-terminal domain (i.e. Met-174), and 2 that are located on the C-terminal domain (i.e. Met-238 and Met-267). The CnBr digestion of un-oxidized yeast 3-PGK[70] results in 4 CnBr fragments with molecular weight ranges from 4.4 to 22 KDa (see Figure 28). The oxidation behavior of Met-238 and Met-267 are expected to be similar and to report on the thermodynamic stability of the C-terminal domain; whereas the oxidation behavior of Met-174 should report on thermodynamic stability of the N-terminal domain. Unfolding of the C-terminal domain at lower Urea concentrations (≤ 1M) results in protection against CnBr

cleavage at Met-238 and 267 resulting in a CnBr fragment that has the size of fragment B, C and D combined (i.e. 23.4 KDa).

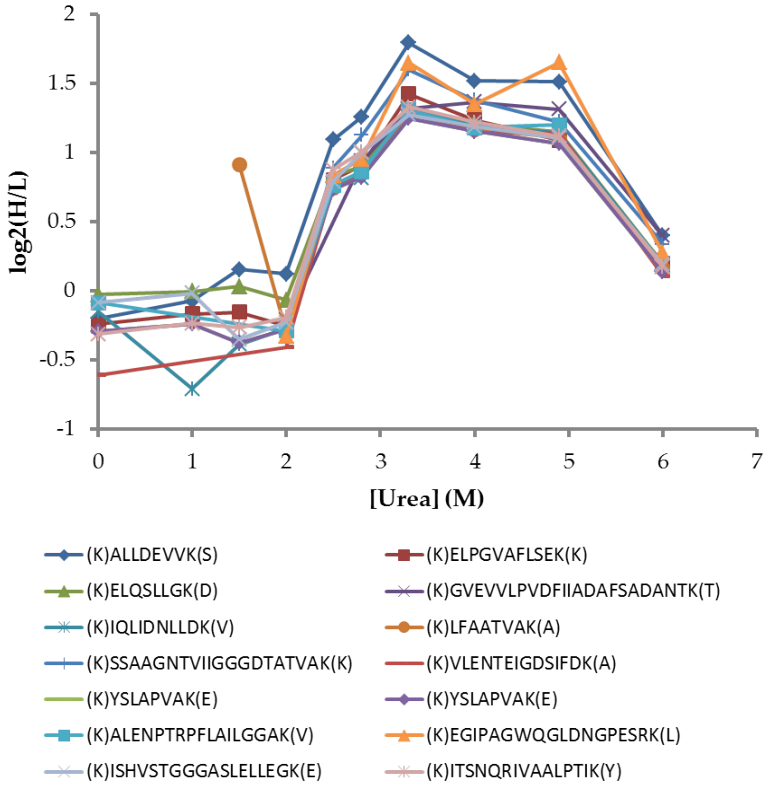

Figure 27: Representative SILAC-SPROX-CnBr data from 3-PGK in gel-based experiment 2.

Unfolding of the N-terminal domain at higher Urea concentrations results in protection against CnBr cleavage at Met-174 and in the appearance of the Full-length protein. The SILAC-SPROX data was collected in the 37-50 KDa molecular weight range. Therefore a gel band excised in this molecular weight range should contain only full-length 3-PGK. The SILAC-SPROX data from all 14 peptides generated from this gel band showed a very reproducible curve with a peak initiating at 2M and continuing on to

more than 6 M Urea. The transition at 2 M urea in this case suggests an averaged $C_{1/2}$ value of both the C-terminal and N-terminal domains. This data also indicates a destabilization. This 3-PGK example demonstrates the power of the gel-based (i.e. the SILAC-SPROX-CnBr) strategy to detect protein-ligand binding that may have been missed in the solution-based SPROX experiments.

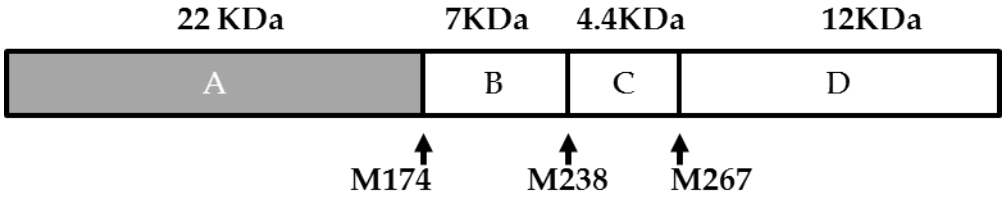

Figure 28: CnBr digestion pattern of the yeast Phosphoglycerate Kinase 1 (3-PGK). Arrows indicate CnBr cleavage sites, upper numbers indicate molecular weight of corresponding CnBr fragments.

5.3.5 ATP binding Properties of Yeast Proteins

5.3.5.1 ATP binding is promiscuous

The solution and gel-based SILAC-SPROX experiments described here and in Chapter 4 identified a total of 140 protein targets of ATP. A fraction of the ATP binding proteins identified in this work, 28%, were previously annotated with GO-term "ATP binding." A large fraction (21%) of the ATP-binding hits identified in this work that were not previously annotated with GO-term "ATP binding", are proteins known to bind nucleoside or nucleotides (e.g., uridine, or GTP) or to bind cofactors with the

adenosine moiety (NAD^+, FAD). The similarity of these ligands to ATP suggests that there is some promiscuity to nucleotide binding proteins.

ATP binding consists of two parts: the phosphate moiety and the adenine based recognition motifs[71]. Among the phosphate binding motifs is the P-loop, a Glycine-rich sequence followed by a conserved Lysine (K) and a Serine (S) or Threonine (T). Interestingly, some GTP-binding proteins such as Elongation factors (i.e. EF-Tu, EF-1α, EF-2 and EF-G) also contain the P-loop[68]. This may explain why many GTP-binding proteins are identified as hits in the ATP-binding experiments. EF-Tu and EF-G are found to be ATP-binding hits in a previous energetics-based study using pulse proteolysis in E. coli[21b]. The corresponding yeast homologs of these proteins, EF-1α and EF-2, were also identified as hits in this SILAC-SPROX ATP-binding study (i.e. in Gel-based experiment 2 and Solution-based experiment 2, respectively). Other GTPase Elongation factors such as EF-1β, EF-3A are also identified as hits (i.e. in Solution-based experiment 2 and Gel-based experiment 2, respectively).

The mode of protein-adenine recognition for ATP is similar to that of other adenine containing cofactors including coenzyme A (CoA) and NAD^+/NADP (Nicotinamide Adenine Dinucleotide Phosphate) [71]. The experimental data from this ATP-binding study also suggests there is a common motif between ATP and NAD^+ as well as FAD binding proteins. These results are also in good agreement with results

obtained from previous ATP-binding study in *E. coli* and Mycobacterium tuberculosis (Mtb) [21a, b, 59].

5.3.5.2 ATP-binding comprises of both weak and tight bindings

In this ATP-binding study, a total of 109 known ATP-binding proteins are effectively assayed (i.e. proteins annotated with GO-term "ATP binding" in the Yeast Genome Database). A total of 37 ATP-binding proteins are detected in this ATP-binding study, which accounts for 30% of the assayed ATP-binding proteins. The other 70% of the assayed ATP-binding proteins presumably have low affinity toward ATP and thus escape from this analysis. This result agrees well with the ATP-binding properties found in Mycobacterium tuberculosis (Mtb)[59]. In this study the proteins from Mtb lysate were incubated with a chemical probe (desthiobiotin-conjugated ATP) in the presence and absence of an ATP inhibitor (ATPγS). If ATP-binding is tight and competitive, the presence of excess ATPγS will out-compete the binding of the chemical probe (i.e. the desthiobiotin-conjugated ATP). In this study, 2 different sets of proteins were identified: those that bind weakly (no difference in presence/absence of ATPγS) and those that bind tightly (significant reduced binding to desthiobiotin-conjugated ATP in presence of ATPγS)[59]. The results from this study suggest a 50:50 ratio between the tight- and weak-ATP binding proteins.

5.3.5.3 Many of ATP-binding Hits Show a Destabilization

Of the 140 ATP-binding hits identified in this study, only 7 proteins show stabilization upon binding to ATP; and 4 of these 7 proteins had some peptide probes showing a destabilization and some peptide probes showing stabilization (see Table 11). The large majority of the ATP-binding hits found in this study showed only a decrease in the thermodynamic stability of proteins upon binding to ATP (see Appendix B).

The findings are consistent with the ATP-binding study in E. coli using the pulse proteolysis approach (see discussion in Chapter 4). In this study, several proteins were found to be destabilized by ATP-binding, e.g. GAPDH, GroEL and Uridine phosphorylase. Interestingly, the GroEL complex was reportedly destabilized by ADP and AMP-PNP; AMP-PNP destabilizes the complex to a lesser extent than ADP (i.e. approximately 10.5 kcal/mol vs. 14.1 kcal/mol)[72]. Apparent destabilization of GroEL by ADP/AMP-PNP, similar to the case of GAPDH, results from dissociation of the 14 mer complex in the presence of these ligands. However, it is noteworthy that the destabilized proteins represented a much smaller fraction of the hits (i.e. 3 out of 30) than in this yeast study.

Many of the destabilized proteins identified in the SILAC-SPROX ATP-binding study are also multimeric proteins. For example, Alcohol dehydrogenase and GAPDH are tetramers; Pyruvate kinase is a dimer, and Aminoacyl tRNA synthetase class II are usually dimers and multi-domain proteins, etc. Apparent destabilization resulting from

dissociation of multimeric proteins may be because of the ligands binding to partially unfolded intermediates and may not report on the change in global thermodynamic stability between the folded and unfolded state of the proteins.

Table 11: Estimated K_d of proteins that have peptides showing stabilization upon binding to ATP. The K_d can only be estimated for peptides that show stabilization and if there are enough data points in the transition regions for re-construction of SPROX curves from SILAC-SPROX data. N/A means there is no calculated K_d available.

Peptides	*Accession #*	*Proteins*	*$\Delta C_{1/2}$*	*Calculated K_d (μM)*	*Experiment*
AGM(ox)TTIVRDLDRPGSK	P14126	60S ribosomal protein L3	Stabilize	N/A	Solution based 1B
TITPM(ox)GGFVHYGEIK	P14126	60S ribosomal protein L3	Destabilize	N/A	Solution based 1B
TITPMGGFVHYGEIK	P14126	60S ribosomal protein L3	Destabilize	N/A	Solution based 1B
(ac)SSSESIRM(ox)VLIGPPGAGK	P07170	Adenylate kinase 1	1.5	204.7	Solution based 1B
(ac)SSSESIRMVLIGPPGAGK	P07170	Adenylate kinase 1	1.2	343.4	Solution based 1B
GTQAPNLQERFHAAHLATGDM(ox)LRSQIAK	P07170	Adenylate kinase 1	1	484.7	Solution based 1B
GTQAPNLQERFHAAHLATGDMLRSQIAK	P07170	Adenylate kinase 1	1	484.7	Solution based 1B
DLLLLDVAPLSLGVGM(ox)QGDM(ox)FGIVVPRNTTVPTIK	P11484	Heat shock protein SSB1	Destabilize	N/A	Solution based 2
M(ox)VNQAEEFK	P40150	Heat shock protein SSB1	1.5	570.1	Solution based 2
NIPM(ox)M(ox)PAGEPVLEAIFEVDANGILK	P40150	Heat shock protein SSB1	-1.75	N/A	Solution based 2
ESEPM(ox)EVDEDDSK	P15705	Heat shock protein STI1	1.5	570.1	Solution based 2
LM(ox)SFPEAIADCNK	P15705	Heat shock protein STI1	Destabilize	N/A	Solution based 2
GVVDSEDLPLNLSREM(ox	P15108	HSP82	-2.25	N/A	Solution-

)LQQNK					based 2
GVVDSEDLPLNLSREMLQ QNK	P15108	HSP82	Destabiliz e	N/A	Solution-based 2
LLDAPAAIRTGQFGWSA NM(ox)ERIM(ox)K	P15108	HSP82	-2	N/A	Solution-based 2
VLEIRDSGIGMTK	P15108	HSP82	Stabilize	N/A	Solution-based 2
TVM(ox)IAAHGNSLRGLV K	P00950	Phosphoglyc erate mutase 1	Stabilize	N/A	Solution_b ased 1A
TVM(ox)IAAHGNSLRGLV K	P00950	Phosphoglyc erate mutase 1	1.75	133.0	Solution_b ased 1B
TVM(ox)IAAHGNSLRGLV K	P00950	Phosphoglyc erate mutase 1	1	1519.2	Solution-based 2
TVMIAAHGNSLRGLVK	P00950	Phosphoglyc erate mutase 1	Stabilize	N/A	Solution_b ased 1A
TVMIAAHGNSLRGLVK	P00950	Phosphoglyc erate mutase 1	1.75	133.0	Solution_b ased 1B

5.3.6 Sensitivity of the SILAC-SPROX protocol

An average protein contains approximately 300 residues and a typical folding domain often contains 60 – 100 residues. An average m value of proteins unfolding in Urea is approximately 0.013 per residue[73], thus an average m value for an average folding domain should be from 0.78 to 1.3 kcal mol^{-1} M^{-1}. Assuming this average m values, a detectable $\Delta C_{1/2}$ of 1 M Urea and the free Ligand concentration of 1 mM, the expected K_d in this first ATP-binding experiment should be approximately in the range of 90 – 400 μM (See equation (12), chapter 1). 1 mM ATP is the effective free ligand concentration utilized in the first ATP-binding experiments (i.e. Solution-based

experiment 1A/B, Gel-based experiment 1). Therefore, ATP binding hits detected in these first experiments are fewer and tighter ATP-binding proteins. For instance, GAPDH, ATP is estimated to bind to an *E. coli* homolog of this protein as tight as 150 μM[64]. Another reported K_d values of dimeric GAPDH to ATP is 450 μM (See discussion in Chapter 4).

One reason for the inability to detect the binding of ATP to nearly 70% of previously annotated ATP binding proteins may be that this fraction of ATP binding proteins have binding affinities that are relatively weak and below the threshold of these analyses. Therefore, in ATP-binding experiment 2 (i.e. solution-based experiment 2 and gel-based experiment 2), the ATP-binding condition and SPROX reaction were adjusted to improve the sensitivity of this assay. First increasing oxidation time maximizes the $C_{1/2}$ shift (See discussion in Histidine HDX chapter 2). In the second ATP-binding experiment, the measurable $C_{1/2}$ shift is in the range of 0.75 M to 2 M Urea. Increasing effective free ligand concentration increases the measurable Kd (i.e. calculated using equation (12)) from 90-400 μM to approximately 600-2700 μM. As a result, the number of ATP-binding hits identified in this second ATP-binding experiment increase from 27 to 99. The number of known ATP-binding proteins increases from 4 to 29. Nevertheless, this fraction only accounts for approximately 30% of known ATP-binding proteins assayed; we hypothesize that the binding of the other known ATP-binding proteins have K_d higher than 2.7 mM. The fraction of detected known ATP-binding proteins in this

work is about half of that in the in E. coli and Mtb studies (both are approximately 70%)[21a, b, 59].

It is noteworthy that the ATP-binding study in *E. coli* and Mtb utilized a slightly different ATP-analog than that employed in the SILAC-SPROX experiments (i.e. ATPγS vs. AMP-PNP). ATPγS significantly reduces the rate of ATP hydrolysis in ATP-binding proteins whereas AMP-PNP is completely non-hydrolysable. Because ATP-binding systems vary greatly in their specificity so what acts as a good analog in one case may be very poor in another[74]. For instance, in a recent study regarding Clamp/Clamp loader proteins in Archaeoglobus fulgidus (*af*RFC), the ability of different ATP analogs (i.e. ATPγS and AMP-PNP) to induce native interaction between *af*RFC and *af*PCNA (Proliferating Cell Nuclear Antigen homologue) was compared with ATP[75]. In this system, the binding of ATP but not the hydrolysis of the nucleotide is enough to induce the *af*RFC-*af*PCNA interaction. The amount of *af*PCNA recruited to *af*RFC by ATP is considered 100%. Whereas 1mM ATPγS was able recruit comparable amount of *af*PCNA to *af*RFC, 1mM AMP-PNP only recruits 1/3 of that by ATP. Increasing this AMP-PNP concentration from 1mM to 10mM increases the amount of *af*PCNA recruited to *af*RFC to comparable to that by ATP. It is unknown why 2 seemingly similar ATP analogs have different effect on ATP dependent protein activity and this can be the reasons for discrepancy between the SILAC-SPROX ATP-binding study and other ATP-binding studies in *E. coli* and Mtb.

5.4 Conclusions and Future Directions

The SILAC-SPROX strategy (including both solution-based and gel-based approaches) successfully assayed 526 proteins in the yeast proteome for ATP binding. A total of 109 of the assayed proteins were known to bind ATP based on GO-term "ATP binding" annotated in the Yeast Genome database. A total of 140 proteins hits were identified as potential ATP-binding targets in yeast, 37 of which have known ATP-binding interactions. An additional 8 proteins detected in these yeast studies have homologs that have been found to interact with ATP in *E. coli*.

The SILAC-SPROX-CnBr approach eliminated the pre-requisite to identify and quantify methionine containing peptides and improved the proteome coverage of solution-based approach by 5-fold. The SILAC-SPROX-CnBr approach not only increases the number of peptides (proteins) assayed in general but it also increased the number of peptide probes per protein. This improvement changes the fundamental basis of current SPROX protocol from a peptide-based quantitation to a protein-based quantitation. However, there are benefits of performing both gel-based and solution-based SILAC-SPROX experiments as the results from both experiments are complementary. Altogether the SILAC-SPROX protocol described here has the potential to assay 1,000 yeast proteins simultaneously for thermodynamic stability changes and protein-ligand interactions.

One unique feature of this SILAC-SPROX analysis is the ability to tune its sensitivity toward the ligand of interest by (i) changing free ligand concentration (ii) changing equilibration time and (iii) changing oxidation time. This flexibility allows the SILAC-SPROX protocol to be applied on a wide variety of ligand classes with wide range of binding affinity (i.e. up to 2-3 mM). This is demonstrated in the ATP-binding study described here, changes in aforementioned conditions in second ATP-binding experiment in solution increased the number of identified protein hits by approximately 5 fold.

The SILAC-SPROX-CnBr approach is also amenable to different gel-based fractionation and quantification protocol. One of which is Difference gel electrophoresis (DIGE)[76]. In this approach, two different biological samples can be labeled with two different fluorescent dyes. This allows combination of the 2 samples and simultaneous comparative analysis between the two on 2-Dimensional gel electrophoresis. In the SILAC-SPROX-CnBr experiment, one can subject the (-) and (+) ligand samples to labeling by 2 different fluorescent dyes preceding digestion by CnBr. The (-) and (+) ligand sample can then be combined and analyzed simultaneously on 2-D gel electrophoresis. In this experiment, each SILAC-SPROX samples from one denaturant concentration is analyzed on a 2-D gel. As described previously, the full-length proteins will appear as the corresponding CnBr fragments appear. The rate of appearance will be different between the (-) and (+) ligand samples resulting in distinguish fluorescence

color at a certain spot on the 2-D gel. Identification of proteins with actual altered H/L ratios from this gel spot will suggest potential binding targets. This approach promises to change the basis of current SPROX protocol from a shotgun proteomic measurement to a targeted approach.

Appendix A

Start AA	sequence	Accession #	Protein Name	Comments
98	(K)ISDDILSVLDSHLIPS ATTGESK(V)	6320304	14-3-3 protein, minor isoform, Bmh2p	No structure
170	(R)AHSSMVGFDLPQRA AGFLLEKELKYFGK(A)	10383781	3-phosphoglycerate kinase, Pgk1p	3pGK, exposed
225	(K)GYLADDIDADSLED IYTSAHEAIR(A)	133003	60S ribosomal protein L5, L1a	3O5H, not enough density to determine orientation of the His
86	(R)YGITHGLTNWAAA YATGLLIAR(R)	133003	60S ribosomal protein L5, L1a	3O5H,Not enough density to determine orientation, part of the loop
179	(K)DLTHVEPPKDLDVI LVAPK(G)	6323387	Acetohydroxyacid reductoisomerasemitochondrial DNA and found in mitochondrial nucleoids	No structure
96	(R)VAPEEHPVLLTEAP MNPK(S)	3328	actin	1YVN,1VAG,Exposed
148	(R)TTGIVLDSGDGVTH VVPIYAGFSLPHAILR(L)	3328	actin	1YVN,part of beta sheet facing to contact with loops regions, seems to be either exposed or partially protected
235	(K)ATDGGAHGVINVS VSEAAIEASTR(Y)	1168350	Alcohol dehydrogenase 1, YADH-1	2HCY, exposed
234	(R)LVNHFIQEFK(R)	6323004	ATP binding protein involved in protein folding and vacuolar import of proteins	No structure
249	(R)IYAFHAEGVRLLK(I)	6320703	Component of CORVET tethering complex, Vps3p	No structure
124	(K)HVVFGEVVDGYDIV K(K)	6320359	Cytoplasmic peptidyl-prolyl cis-trans isomerase (cyclophilin),Cpr1p	1IST, exposed, loop region
481	(R)EMGEEVYHAYTEVK (A)	6325196	Cytosolic aldehyde dehydrogenase	No structure
63	(K)TVDESNVDPDELLF DTELADEDLLTHDAR(	6321724	Cytosolic ribosome-associated chaperone, Zuo1p	No structure

	D)			
786	(R)QATGGQAFPQMVFDHWSTLGSDPLDPTSK(A)	6320593	Elongation factor 2 (EF-2)	1N0V
33	(R)SIVPSGASTGVHEALEMR(D)	119336	Enolase 1	3ENL, exposed
289	(K)RYPIVSIEDPFAEDDWEAWSHFFK(T)	119336	Enolase 1	3ENL, exposed
498	(K)ATHILDFGPGGASGLGVLTHR(N)	3694	fatty acid synthetase subunit beta	2PFF, loop, exposed
11	(K)TGVIVGEDVHNLFTYAK(E)	6322790	Fructose 1,6-bisphosphate aldolase	No structure
133	(K)QSDLVHIVNSWQALTDDLSR(E)	6319492	Galactose-1-phosphate uridyl transferase	No structure
376	(R)GNVFTDYSTGSILFGEPATNAQHSFFQLVHQGTK(L)	6319673	Glycolytic enzyme phosphoglucose isomerase, Pgi1p	No structure
11	(R)GTLEGHNGWVTSLATSAGQPNLLLSASR(D)	6323763	G-protein beta subunit, Asc1p	3RFG, 3RFH,seems exposed part of the loop but facing inward
456	(R)THLMSPAMAAAAGIAGHFVDIR(E)	6321429	Isopropylmalate isomerase	No structure
571	(R)HFVELPR(L)	20428568	major coat protein, Pdc1p	1M1C Exposed
244	(R)HALTVQDADEWIEGDRTDDQFRPPSSK(V)	20428568	major coat protein,Pdc1p	1M1C, exposed
244	(R)HALTVQDADEWIEGDR(T)	20428568	major coat protein,Pdc1p	1M1C, exposed
485	(K)AQYNEIQGWDHLSLLPTFGAK(D)	6323073	Major of three pyruvate decarboxylase isozymes	1PYD,On the loop oriented inward
110	(K)QLLLHHTLGNGDFTVFHR(M)	6321524	Minor isoform of pyruvate decarboxylase, Pdc6p	No structure
299	(K)IHPIGELSSEEEEMLQK(C)	6322765	Mitochondrial malate dehydrogenase	No structure

112	(R)IAEWADITNAHGVVGPGIVSGLK(Q)	6320814	Orotidine-5'-phosphate (OMP) decarboxylase, Ura3p	1DQW,kinda partially protected, part of beta sheet but facing outward to the pocket and maybe in contact with the alpha chains
272	(K)VGYTLPSHIISTSDVTR(I)	6320215	Protein component of the large (60S) ribosomal subunit	1S1I,2WWA, sequence in structure didn't cover this His
143	(R)VDFKNPHDIIEGINAGEIEIPEN(-)	6320625	Protein component of the large (60S) ribosomal subunit	3O5H,Not included in the structure
57	(K)ALAIFVPVPSLAGFHK(V)	6324670	Protein component of the small (40S) ribosomal subunit	No structure
90	(R)SVAVPVDILDHDNNYELK(V)	6319546	Small heat shock protein (sHSP) with chaperone activity	No structure
141	(R)ENTEGEFSGLEHESVPGVVESLK(V)	6324291	Subunit of mitochondrial NAD(+)-dependent isocitrate dehydrogenase	3BLX, seems to be protected, low intensity
157	(K)ETNPGTDVTVSSVESVLAHL(-)	6323138	Thiol-specific peroxiredoxin, Ahp1p	No structure
405	(R)NTAAWHPVIENLVGTVDDDSLVSIYKPYTEESE(-)	6324804	Transcriptional repressor involved in regulation of meiosis and silencing,Wtm1p and Rnr4p	No structure
145	(K)AIEMEGLTWGAHQFIPIGFGIK(K)	6319315	Translation elongation factor 1 beta	1F60, exposed
89	(R)KEGGLGPINIPLLADTNHSLSR(D)	6323613	Ubiquitous housekeeping thioredoxin peroxidase	No structure
71	(R)AIPSALSYVGEDEYHGGQALQQLIR(N)	487960	Yhr064cp	No structure

Appendix B

Peptide sequence	Accession #	Protein name	Experiment	Known Ligand
M(ox)TDYFACGDDDVK	P22146	1,3-beta-glucanosyltransferase GAS1	Solution-based Exp.2	N/A
LAM(ox)VVPGSGLVK	P07264	3-isopropylmalate dehydratase	Gel-based Exp. 2	4Fe-4S, Fe, Fe-S, Metal ion
LAM(ox)VVPGSGLVK	P07264	3-isopropylmalate dehydratase	Solution-based Exp.2	4Fe-4S, Fe, Fe-S, Metal ion
LLEM(ox)STEDFVK	Q01855	40S ribosomal protein S15	Gel-based Exp. 2	Ribosomal Protein
LLEM(ox)STEDFVK	Q01855	40S ribosomal protein S15	Solution-based Exp.1B	Ribosomal Protein
LLEM(ox)STEDFVK	Q01855	40S ribosomal protein S15	Solution-based Exp.2	Ribosomal Protein
RAGELTQEELERIVQIM(ox)QNPTHYK	P0CX55	40S ribosomal protein S18-A;40S ribosomal protein S18-B	Solution-based Exp.1B	RNA
IGIVEISPK	P07280	40S ribosomal protein S19-A	Gel-based Exp. 2	Ribosomal Protein
LEVPGYVDIVK	P07280	40S ribosomal protein S19-A	Gel-based Exp. 2	Ribosomal Protein
TSSGNEM(ox)PPQDAEGWFYK	P07281	40S ribosomal protein S19-A;40S ribosomal protein S19-B	Solution-based Exp.1B	Ribosomal Protein
TSSGNEM(ox)PPQDAEGWFYK	P07280	40S ribosomal protein S19-A;40S ribosomal protein S19-B	Solution-based Exp.2	Ribosomal Protein
EVQGSTLAQLTSK	P33442	40S ribosomal protein S1-A	Gel-based Exp. 2	Ribosomal Protein
FDVGALM(ox)ALHGEGSGEEK	P33442	40S ribosomal protein S1-A	Gel-based Exp. 2	Ribosomal Protein
LIPEVINK	P33442	40S ribosomal protein S1-A	Gel-based Exp. 2	Ribosomal Protein
NLLTNFHGM(ox)DFTTDK	P33442	40S ribosomal protein S1-A	Gel-based Exp. 2	Ribosomal Protein
VTGFKDEVLETV	P33442	40S ribosomal protein S1-A	Gel-based Exp. 2	Ribosomal Protein
VSGFKDEVLETV	P23248	40S ribosomal protein S1-B	Gel-based Exp. 2	Ribosomal Protein

FDVGALM(ox)ALHGEGSGEEK	P23248	40S ribosomal protein S1-B;40S ribosomal protein S1-A	Solution-based Exp.2	Ribosomal Protein
NLLTNFHGM(ox)DFTTDK	P23248	40S ribosomal protein S1-B;40S ribosomal protein S1-A	Solution-based Exp.2	Ribosomal Protein
NLLTNFHGMDFTTDK	P23248	40S ribosomal protein S1-B;40S ribosomal protein S1-A	Solution-based Exp.2	Ribosomal Protein
EFQIIDTLLPGLQDEVM(ox)NIKPVQK	P25443	40S ribosomal protein S2	Solution-based Exp.2	Ribosomal Protein
ITTIEEIFLHSLPVK	P25443	40S ribosomal protein S2	Gel-based Exp. 2	Ribosomal Protein
IVVQLNGRLNK	P0C0W1	40S ribosomal protein S22-A	Gel-based Exp. 2	Ribosomal Protein
ALPDAVTIIEPK	P05750	40S ribosomal protein S3	Gel-based Exp. 1	RNA
ALPDAVTIIEPK	P05750	40S ribosomal protein S3	Gel-based Exp. 2	RNA
ALPDAVTIIEPK	P05750	40S ribosomal protein S3	Gel-based Exp. 2	RNA
ALPDAVTIIEPKEEEPILAPSVK	P05750	40S ribosomal protein S3	Gel-based Exp. 2	RNA
EEEPILAPSVK	P05750	40S ribosomal protein S3	Gel-based Exp. 1	RNA
EEEPILAPSVK	P05750	40S ribosomal protein S3	Gel-based Exp. 2	RNA
EEEPILAPSVK	P05750	40S ribosomal protein S3	Gel-based Exp. 2	RNA
YAPGTIVLYAERVQDRGLSAVAQAES M(ox)K	P05750	40S ribosomal protein S3	Solution-based Exp.2	RNA
TIAETLAEELINAAK	P26783	40S ribosomal protein S5	Gel-based Exp. 1	RNA
WSFEEVEVK	P26783	40S ribosomal protein S5	Gel-based Exp. 1	RNA
VLEDM(ox)VFPTEIVGK	P48164	40S ribosomal protein S7-B	Gel-based Exp. 1	Ribosomal Protein
VLEDMVFPTEIVGK	P48164	40S ribosomal protein S7-B	Solution-based Exp.1B	Ribosomal Protein
QIVNIPSFM(ox)VRLDSEK	P05755	40S ribosomal protein S9-A;40S ribosomal protein S9-B	Solution-based Exp.2	RNA
AYTYFGEQSNLPK	P05694	5-methyltetrahydropteroyltriglutamate--homocysteine met	Gel-based Exp. 2	Metal, Zinc

FINSEIEK	P05694	5-methyltetrahydropteroyltriglutamate--homocysteine met	Gel-based Exp. 2	Metal, Zinc
GTISAEEYEK	P05694	5-methyltetrahydropteroyltriglutamate--homocysteine met	Gel-based Exp. 2	Metal, Zinc
LASAGATEVQIDEPVLVLDLPANAQAAIK	P05694	5-methyltetrahydropteroyltriglutamate--homocysteine met	Gel-based Exp. 2	Metal, Zinc
LDEVVVITK	P05694	5-methyltetrahydropteroyltriglutamate--homocysteine met	Gel-based Exp. 1	Metal, Zinc
LDEVVVITK	P05694	5-methyltetrahydropteroyltriglutamate--homocysteine met	Gel-based Exp. 2	Metal, Zinc
M(ox)STRAAPFEQRLPEQQK	P05694	5-methyltetrahydropteroyltriglutamate--homocysteine met	Gel-based Exp. 2	Metal, Zinc
NVSGQDVAAALEANAK	P05694	5-methyltetrahydropteroyltriglutamate--homocysteine met	Gel-based Exp. 2	Metal, Zinc
VFNLPLFPTTTIGSFPQTK	P05694	5-methyltetrahydropteroyltriglutamate--homocysteine met	Gel-based Exp. 2	Metal, Zinc
ELRGRAVVLM(ox)GK	P05317	60S acidic ribosomal protein P0	Solution-based Exp.1B	Ribosomal Protein
LREYLEEYK	P05317	60S acidic ribosomal protein P0	Gel-based Exp. 1	Ribosomal Protein
NTM(ox)VRRAIRGFLSDLPDFEK	P05317	60S acidic ribosomal protein P0	Solution-based Exp.1B	Ribosomal Protein
NVIVSNRVAAPARAGAVAPEDIWVRAVNTGM(ox)EPGK	P05317	60S acidic ribosomal protein P0	Solution-based Exp.1B	Ribosomal Protein
NVIVSNRVAAPARAGAVAPEDIWVRAVNTGMEPGK	P05317	60S acidic ribosomal protein P0	Solution-based Exp.1B	Ribosomal Protein
SLFVVGVDNVSSQQM(ox)HEVRK	P05317	60S acidic ribosomal protein P0	Gel-based Exp. 1	Ribosomal Protein

SLFVVGVDNVSSQQM(ox)HEVRK	P05317	60S acidic ribosomal protein P0	Solution-based Exp.1B	Ribosomal Protein
SLFVVGVDNVSSQQMHEVRK	P05317	60S acidic ribosomal protein P0	Solution-based Exp.1B	Ribosomal Protein
RNFLETVELQVGLK	P53030	60S ribosomal protein L1	Gel-based Exp. 2	Ribosomal Protein
IGPLGLSPK	P17079	60S ribosomal protein L12	Gel-based Exp. 2	RNA
HSGNIQLDEIIEIARQM(ox)RDK	P0CX53	60S ribosomal protein L12-A;60S ribosomal protein L12-B	Solution-based Exp.1B	RNA
HSGNIQLDEIIEIARQMRDK	P0CX53	60S ribosomal protein L12-A;60S ribosomal protein L12-B	Solution-based Exp.1B	RNA
IVVVRAEALNISGEFFRNK	P26785	60S ribosomal protein L16-B	Gel-based Exp. 1	RNA
SCGVDAM(ox)SVDDLK	P0CX43	60S ribosomal protein L1-A;60S ribosomal protein L1-B	Solution-based Exp.2	Ribosomal Protein
EYQVIGRRLPTESVPEPK	P0C2I0	60S ribosomal protein L20	Gel-based Exp. 2	Ribosomal Protein
AQGVAVQLK	Q02753	60S ribosomal protein L21-A	Gel-based Exp. 2	Ribosomal Protein
SSVGVIINK	Q02753	60S ribosomal protein L21-A	Gel-based Exp. 2	Ribosomal Protein
TGVVYNVTK	Q02753	60S ribosomal protein L21-A	Gel-based Exp. 2	Ribosomal Protein
GSGSRLNRLPAASLGDM(ox)VM(ox)ATVK	P0CX41	60S ribosomal protein L23-A;60S ribosomal protein L23-B	Solution-based Exp.2	Ribosomal Protein
M(ox)KVEVDSFSGAK	P24000	60S ribosomal protein L24-B	Solution-based Exp.1B	RNA
KVEDGNILVFQVSM(ox)K	P04456	60S ribosomal protein L25	Solution-based Exp.1B	RNA
VIEQPITSETAMK	P04456	60S ribosomal protein L25	Solution-based Exp.2	RNA
ETAPVIDTLAAGYGK	P02406	60S ribosomal protein L28	Gel-based Exp. 2	Ribosomal Protein
LWTLIPEDK	P02406	60S ribosomal protein L28	Gel-based Exp. 2	Ribosomal Protein
AGM(ox)TTIVRDLDRPGSK	P14126	60S ribosomal protein L3	Gel-based Exp. 2	Ribosomal Protein
AGM(ox)TTIVRDLDRPGSK	P14126	60S ribosomal protein L3	Solution-based Exp.1B	Ribosomal Protein
AHLAEIQLNGGSISEK	P14126	60S ribosomal protein L3	Gel-based Exp. 2	Ribosomal Protein
ALEEVSLK	P14126	60S ribosomal protein L3	Gel-based Exp. 2	Ribosomal Protein

DDRSKPVALTSFLGYK	P14126	60S ribosomal protein L3	Gel-based Exp. 2	Ribosomal Protein
IYRVGKGDDEANGATSFDRTK	P14126	60S ribosomal protein L3	Gel-based Exp. 2	Ribosomal Protein
TITPM(ox)GGFVHYGEIK	P14126	60S ribosomal protein L3	Gel-based Exp. 2	Ribosomal Protein
TITPM(ox)GGFVHYGEIK	P14126	60S ribosomal protein L3	Solution-based Exp.1B	Ribosomal Protein
TITPM(ox)GGFVHYGEIK	P14126	60S ribosomal protein L3	Solution-based Exp.2	Ribosomal Protein
TITPMGGFVHYGEIK	P14126	60S ribosomal protein L3	Solution-based Exp.1B	Ribosomal Protein
TVAVDSVFEQNEM(ox)IDAIAVTK	P14126	60S ribosomal protein L3	Gel-based Exp. 2	Ribosomal Protein
VDWAREHFEK	P14126	60S ribosomal protein L3	Gel-based Exp. 2	Ribosomal Protein
LIIIAANTPVLRK	P14120	60S ribosomal protein L30	Gel-based Exp. 2	Ribosomal Protein
LHM(ox)GTDDVRLAPELNQAIWK	P0C2H8	60S ribosomal protein L31-A	Solution-based Exp.1B	Ribosomal Protein
LHM(ox)GTDDVRLAPELNQAIWK	P0C2H8	60S ribosomal protein L31-A	Solution-based Exp.2	Ribosomal Protein
M(ox)GVPYAIVK	P17076	60S ribosomal protein L8-A	Solution-based Exp.2	Ribosomal Protein
M(ox)RYVYAHFPINVNIVEK	P05738	60S ribosomal protein L9-A;60S ribosomal protein L9-B	Solution-based Exp.1B	Ribosomal Protein
SLVDNM(ox)ITGVTK	P05738	60S ribosomal protein L9-A;60S ribosomal protein L9-B	Solution-based Exp.1B	Ribosomal Protein
SLVDNM(ox)ITGVTK	P05738	60S ribosomal protein L9-A;60S ribosomal protein L9-B	Solution-based Exp.2	Ribosomal Protein
IIRM(ox)PLVESVK	P16861	6-phosphofructokinase subunit alpha	Solution-based Exp.2	ATP
DLLQTM(ox)FPVDFIHEGK	Q00955	Acetyl-CoA carboxylase;Biotin carboxylase	Solution-based Exp.2	ATP, Biotin, Mn, Metal ion
TTTDHISM(ox)AGPWLK	P19414	Aconitate hydratase, mitochondrial	Solution-based Exp.2	4Fe-4S, Fe, Fe-S, Metal ion, ATP*
EITALAPSSMK	P60010	Actin	Solution-based Exp.1B	ATP
M(ox)AIFDELLQM(ox)PETK	P47143	Adenosine kinase	Solution-based Exp.2	ATP, Mg
IADISLAAFGRK	P39954	Adenosylhomocysteinase	Gel-based Exp. 2	NAD
LNLILDDGGDLTTLVHEK	P39954	Adenosylhomocysteinase	Gel-based Exp. 2	NAD

TGPFEVGVHVLPK	P39954	Adenosylhomocysteinase	Gel-based Exp. 2	NAD
VPAINVNDSVTK	P39954	Adenosylhomocysteinase	Gel-based Exp. 2	NAD
VPAINVNDSVTK	P39954	Adenosylhomocysteinase	Gel-based Exp. 2	NAD
VQSEYLGIPEEGPFK	P39954	Adenosylhomocysteinase	Gel-based Exp. 2	NAD
(ac)SSSESIRM(ox)VLIGPPGAGK	P07170	Adenylate kinase 1	Solution-based Exp.1B	ATP
(ac)SSSESIRMVLIGPPGAGK	P07170	Adenylate kinase 1	Solution-based Exp.1B	ATP
GTQAPNLQERFHAAHLATGDM(ox)LRSQIAK	P07170	Adenylate kinase 1	Solution-based Exp.1B	ATP
DGM(ox)DFADAM(ox)AQSTK	P17555	Adenylyl cyclase-associated protein	Solution-based Exp.2	Actin
NPELRQSSTVSSTGSK	P17555	Adenylyl cyclase-associated protein	Gel-based Exp. 2	Actin
ESNFLIDFLM(ox)GGVSAAVAK	P18239	ADP,ATP carrier protein 2	Solution-based Exp.2	ATP, ADP
LLIQNQDEM(ox)LK	P18239	ADP,ATP carrier protein 2	Solution-based Exp.2	ATP, ADP
AM(ox)GYRVLGIDGGEGK	P00330	Alcohol dehydrogenase 1	Gel-based Exp. 2	NAD, Metal, Zinc
AM(ox)GYRVLGIDGGEGK	P00330	Alcohol dehydrogenase 1	Gel-based Exp. 2	NAD, Metal, Zinc
AM(ox)GYRVLGIDGGEGK	P00330	Alcohol dehydrogenase 1	Solution-based Exp.2	NAD, Metal, Zinc
ANELLINVK	P00330	Alcohol dehydrogenase 1	Gel-based Exp. 2	NAD, Metal, Zinc
ANELLINVK	P00330	Alcohol dehydrogenase 1	Gel-based Exp. 2	NAD, Metal, Zinc
DIVGAVLK	P00330	Alcohol dehydrogenase 1	Gel-based Exp. 2	NAD, Metal, Zinc
DIVGAVLK	P00330	Alcohol dehydrogenase 1	Gel-based Exp. 2	NAD, Metal, Zinc
GQIVGRYVVDTSK	P00330	Alcohol dehydrogenase 1	Gel-based Exp. 2	NAD, Metal, Zinc
LPLVGGHEGAGVVVGM(ox)GENVK	P00330	Alcohol dehydrogenase 1	Gel-based Exp. 2	NAD, Metal, Zinc
VVGLSTLPEIYEK	P00330	Alcohol dehydrogenase 1	Gel-based Exp. 2	NAD, Metal, Zinc
LPLVGGHEGAGVVVGM(ox)GENVK	P00330	Alcohol dehydrogenase 1;Alcohol dehydrogenase 2	Solution-based Exp.2	NAD, Metal, Zinc
LPLVGGHEGAGVVVGMGENVK	P00330	Alcohol dehydrogenase 1;Alcohol dehydrogenase 2	Solution-based Exp.2	NAD, Metal, Zinc
SLQDIIAILGM(ox)DELSEQDK	P00830	ATP synthase subunit beta,	Solution-based Exp.2	ATP

		mitochondrial		
AELINNLGTIAK	P15108	ATP-dependent molecular chaperone HSC82	Gel-based Exp. 2	ATP
AELINNLGTIAK	P15108	ATP-dependent molecular chaperone HSC82	Gel-based Exp. 2	ATP
GVVDSEDLPLNLSREM(ox)LQQNK	P15108	ATP-dependent molecular chaperone HSC82	Gel-based Exp. 2	ATP
GVVDSEDLPLNLSREM(ox)LQQNK	P15108	ATP-dependent molecular chaperone HSC82	Gel-based Exp. 2	ATP
LGVHEDTQNRAALAK	P15108	ATP-dependent molecular chaperone HSC82	Gel-based Exp. 2	ATP
LGVHEDTQNRAALAK	P15108	ATP-dependent molecular chaperone HSC82	Gel-based Exp. 2	ATP
VLEIRDSGIGM(ox)TK	P15108	ATP-dependent molecular chaperone HSC82	Gel-based Exp. 2	ATP
VLEIRDSGIGM(ox)TK	P15108	ATP-dependent molecular chaperone HSC82	Gel-based Exp. 2	ATP
GVVDSEDLPLNLSREM(ox)LQQNK	P15108	ATP-dependent molecular chaperone HSP82;ATP-dependent m	Solution-based Exp.2	ATP
GVVDSEDLPLNLSREMLQQNK	P15108	ATP-dependent molecular chaperone HSP82;ATP-dependent m	Solution-based Exp.2	ATP
LLDAPAAIRTGQFGWSANM(ox)ERIM(ox)K	P15108	ATP-dependent molecular chaperone HSP82;ATP-dependent m	Solution-based Exp.2	ATP
VLEIRDSGIGMTK	P15108	ATP-dependent molecular chaperone HSP82;ATP-dependent m	Solution-based Exp.2	ATP
APQALM(ox)LAPTRELALQIQK	P10081	ATP-dependent RNA helicase eIF4A	Solution-based Exp.1B	ATP
FTVSAIYSDLPQQERDTIM(ox)K	P10081	ATP-dependent RNA helicase eIF4A	Solution-based Exp.2	ATP
GVAINFVTNEDVGAM(ox)RELEK	P10081	ATP-dependent RNA helicase eIF4A	Solution-based Exp.2	ATP
M(ox)FILDEADEM(ox)LSSGFK	P10081	ATP-dependent RNA helicase eIF4A	Solution-based Exp.2	ATP
SPTPQ(de)EILQ(de)RFQ(de)KQ(de)	Q06337	Chromatin modification-related protein EAF1	Gel-based Exp. 1	DNA
LEDNM(ox)NLEAK	REV__P46677	CON__P0C1U8	Solution-based Exp.2	N/A

MVDLGIQPPPISTPNLSLPPVILSRFGSDPSK	P43616	Cys-Gly metallodipeptidase DUG1	Solution-based Exp.2	Mn, Metal ion, Zinc
AM(ox)QFNTIGVSDGISM(ox)GTK	P39522	Dihydroxy-acid dehydratase, mitochondrial	Solution-based Exp.2	4Fe-4S
YLYENNM(ox)LHGNTM(ox)TVTGDTLAERAK	P39522	Dihydroxy-acid dehydratase, mitochondrial	Solution-based Exp.1B	4Fe-4S
YVM(ox)ADLINVGGTQSVIK	P39522	Dihydroxy-acid dehydratase, mitochondrial	Solution-based Exp.2	4Fe-4S
SVPTACNSYGGM(ox)YK	P39976	D-lactate dehydrogenase [cytochrome] 3	Solution-based Exp.2	FAD, Flavoprotein
IGGIGTVPVGRVETGVIKPGM(ox)VVTFAPAGVTTEVK	P02994	Elongation factor 1-alpha	Gel-based Exp. 2	GTP, ATP (EF-Tu)
IGGIGTVPVGRVETGVIKPGM(ox)VVTFAPAGVTTEVK	P02994	Elongation factor 1-alpha	Gel-based Exp. 2	GTP, ATP (EF-Tu)
SVEM(ox)HHEQLEQGVPGDNVGFNVK	P02994	Elongation factor 1-alpha	Gel-based Exp. 2	GTP, ATP (EF-Tu)
WDESRFQEIVK	P02994	Elongation factor 1-alpha	Gel-based Exp. 2	GTP, ATP (EF-Tu)
WDESRFQEIVK	P02994	Elongation factor 1-alpha	Gel-based Exp. 2	GTP, ATP (EF-Tu)
WDESRFQEIVK	P02994	Elongation factor 1-alpha	Gel-based Exp. 2	GTP, ATP (EF-Tu)
YAWVLDK	P02994	Elongation factor 1-alpha	Gel-based Exp. 2	GTP, ATP (EF-Tu)
YAWVLDK	P02994	Elongation factor 1-alpha	Gel-based Exp. 2	GTP, ATP (EF-Tu)
YQVTVIDAPGHRDFIK	P02994	Elongation factor 1-alpha	Gel-based Exp. 2	GTP, ATP (EF-Tu)
AIEM(ox)EGLTWGAHQFIPIGFGIK	P32471	Elongation factor 1-beta	Solution-based Exp.2	GTP
SIVTLDVKPWDDETNLEEM(ox)VANVK	P32471	Elongation factor 1-beta	Solution-based Exp.2	GTP
VSLDDLQQSIEEDEDHVQSTDIAAM(ox)QK	P32471	Elongation factor 1-beta	Solution-based Exp.2	GTP
ADLMLYVSK	P32324	Elongation factor 2	Solution-based Exp.2	GTP, ATP* (EF-G)
M(ox)M(ox)DRLWGDSFFNPK	P32324	Elongation factor 2	Solution-based Exp.2	GTP, ATP* (EF-G)
STAISLYSEM(ox)SDEDVK	P32324	Elongation factor 2	Solution-based Exp.2	GTP, ATP* (EF-G)

VAFTVDQMRSLMDK	P32324	Elongation factor 2	Solution-based Exp.2	GTP, ATP* (EF-G)
AYEELSNTDLEFKFPEPGYLEGVK	P16521	Elongation factor 3A	Gel-based Exp. 2	ATP, GTP
EALASGQFRPLTRK	P16521	Elongation factor 3A	Gel-based Exp. 2	ATP, GTP
GNFTEFVK	P16521	Elongation factor 3A	Gel-based Exp. 2	ATP, GTP
LVEDPQVIAPFLGK	P16521	Elongation factor 3A	Gel-based Exp. 2	ATP, GTP
NLTEEVWAVK	P16521	Elongation factor 3A	Gel-based Exp. 2	ATP, GTP
AADALLLK	P00924	Enolase 1	Gel-based Exp. 2	Mg, Metal ion
NVNDVIAPAFVK	P00924	Enolase 1	Gel-based Exp. 2	Mg, Metal ion
RYGASAGNVGDEGGVAPNIQTAEEALDLIVDAIK	P00924	Enolase 1	Gel-based Exp. 2	Mg, Metal ion
VNQIGTLSESIK	P00924	Enolase 1	Gel-based Exp. 2	Mg, Metal ion
GVFRSIVPSGASTGVHEALEM(ox)RDEDK	P00925	Enolase 2	Solution-based Exp.1B	Mg, Metal
GVFRSIVPSGASTGVHEALEM(ox)RDEDK	P00925	Enolase 2	Solution-based Exp.2	Mg, Metal
GVFRSIVPSGASTGVHEALEMRDEDK	P00925	Enolase 2	Solution-based Exp.1B	Mg, Metal
GVMNAVNNVNNVIAAAFVK	P00925	Enolase 2	Solution-based Exp.2	Mg, Metal
LGANAILGVSMAAARAAAAEK	P00925	Enolase 2	Solution-based Exp.2	Mg, Metal
NVPLYQHLADLSK	P00925	Enolase 2	Gel-based Exp. 2	Mg, Metal
NVPLYQHLADLSK	P00925	Enolase 2	Gel-based Exp. 2	Mg, Metal
TAGIQIVADDLTVTNPARIATAIEK	P00925	Enolase 2	Gel-based Exp. 2	Mg, Metal
TFAEAMRIGSEVYHNLK	P00925	Enolase 2	Solution-based Exp.2	Mg, Metal
WLTGVELADM(ox)YHSLM(ox)K	P00925	Enolase 2	Solution-based Exp.2	Mg, Metal
TDVVFM(ox)PVSGYSGANLK	P05453	Eukaryotic peptide chain release factor GTP-binding sub	Solution-based Exp.2	GTP
ATGQEM(ox)DVVSEEPLIEWLAANYK	P12385	Eukaryotic peptide chain release factor subunit 1	Solution-based Exp.2	GTP, TPA-1
M(ox)DLTSFLNDDTFGSSWAEEDVDLNK	P34167	Eukaryotic translation initiation factor 4B	Solution-based Exp.2	RNA

KLEDLSPSTHNM(ox)EVPVVK	P23301	Eukaryotic translation initiation factor 5A-1	Gel-based Exp. 1	RNA*
LEDLSPSTHNM(ox)EVPVVK	P23301	Eukaryotic translation initiation factor 5A-1	Solution-based Exp.2	RNA*
RNEYQLLDIDDGFLSLM(ox)NM(ox)DG DTK	P23301	Eukaryotic translation initiation factor 5A-1	Solution-based Exp.1B	RNA*
RNEYQLLDIDDGFLSLM(ox)NM(ox)DG DTK	P23301	Eukaryotic translation initiation factor 5A-1	Solution-based Exp.2	RNA*
VHLVAIDIFTGK	P23301	Eukaryotic translation initiation factor 5A-1	Gel-based Exp. 1	RNA*
APEGELGDSLQTAFDEGK	P23301	Eukaryotic translation initiation factor 5A-2	Gel-based Exp. 2	RNA*
EM(ox)IQEVIVEEDLEPFEASK	P19097	Fatty acid synthase subunit alpha; Acyl carrier;3-oxoacy	Solution-based Exp.2	NAD, NADP, Mg, Metal ion, phosphopantethein
FFHNGM(ox)IYNK	P19097	Fatty acid synthase subunit alpha;Acyl carrier;3-oxoacy	Solution-based Exp.2	NAD, NADP, Mg, Metal ion, phosphopantethein
FM(ox)EYHISNTDETK	P19097	Fatty acid synthase subunit alpha;Acyl carrier;3-oxoacy	Solution-based Exp.2	NAD, NADP, Mg, Metal ion, phosphopantethein
SPVM(ox)ADLNGGLQFVPELK	P19097	Fatty acid synthase subunit alpha;Acyl carrier;3-oxoacy	Solution-based Exp.2	NAD, NADP, Mg, Metal ion, phosphopantethein
APSGLDQSRIPFSERK	P07149	Fatty acid synthase subunit beta	Gel-based Exp. 2	NAD, NADP
LLQENDTTLVK	P07149	Fatty acid synthase subunit beta	Gel-based Exp. 2	NAD, NADP
GLLSATQFTQPALTLM(ox)EK	P07149	Fatty acid synthase subunit beta;3-hydroxypalmitoyl-[ac	Solution-based Exp.2	NAD, NADP
NYITARIM(ox)AK	P07149	Fatty acid synthase subunit beta;3-	Solution-based Exp.2	NAD, NADP

		hydroxypalmitoyl-[ac		
EDLYTKPEQVYNVYK	P14540	Fructose-bisphosphate aldolase	Gel-based Exp. 2	Metal, Zinc
EHGEPLFSSHM(ox)LDLSEETDEENIST CVK	P14540	Fructose-bisphosphate aldolase	Solution-based Exp.2	Metal, Zinc
GISNEGQNASIK	P14540	Fructose-bisphosphate aldolase	Gel-based Exp. 2	Metal, Zinc
FEVM(ox)LQGIK	P54885	Gamma-glutamyl phosphate reductase	Solution-based Exp.2	NADP
FLM(ox)DGTTYLPSNGGIVQRDYK	Q12680	Glutamate synthase [NADH]	Solution-based Exp.2	3Fe-4S, FAD, FMN, Flavoprotein, Fe, Metal, NAD
DWRGGRTASGNIIPSSTGAAK	P00360	Glyceraldehyde-3-phosphate dehydrogenase 1	Gel-based Exp. 2	NAD, ATP
DWRGGRTASGNIIPSSTGAAK	P00360	Glyceraldehyde-3-phosphate dehydrogenase 1	Gel-based Exp. 2	NAD, ATP
IATYQERDPANLPWGSLK	P00360	Glyceraldehyde-3-phosphate dehydrogenase 1	Gel-based Exp. 2	NAD, ATP
LTGM(ox)AFRVPTVDVSVVDLTVK	P00360	Glyceraldehyde-3-phosphate dehydrogenase 1	Gel-based Exp. 2	NAD, ATP
LTGM(ox)AFRVPTVDVSVVDLTVK	P00360	Glyceraldehyde-3-phosphate dehydrogenase 1	Gel-based Exp. 2	NAD, ATP
VLPELQGK	P00360	Glycer 3-phosphate dehyd	Gel-based Exp. 2	NAD, ATP
VLPELQGK	P00360	Glyceraldehyde-3-phosphate dehydrogenase 1	Gel-based Exp. 2	NAD, ATP
VLPELQGK	P00360	Glyceraldehyde-3-phosphate dehydrogenase 1	Gel-based Exp. 2	NAD, ATP
VVITAPSSSAPM(ox)FVVGVNHTK	P00360	Glyceraldehyde-3-phosphate dehydrogenase 1	Solution-based Exp.1B	NAD, ATP
VVITAPSSSAPM(ox)FVVGVNHTK	P00360	Glyceraldehyde-3-phosphate dehydrogenase 1	Solution-based Exp.2	NAD, ATP
VVITAPSSSAPMFVVGVNHTK	P00360	Glyceraldehyde-3-phosphate	Solution-based Exp.1B	NAD, ATP

		dehydrogenase 1		
LVSWYDNEYGYSTRVVDLVEHVAK	P00358	Glyceraldehyde-3-phosphate dehydrogenase 2	Gel-based Exp. 2	NAD, ATP
NVEVVALNDPFISNDYSAYM(ox)FK	P00358	Glyceraldehyde-3-phosphate dehydrogenase 2	Gel-based Exp. 2	NAD, ATP
NVEVVALNDPFISNDYSAYM(ox)FK	P00358	Glyceraldehyde-3-phosphate dehydrogenase 2	Solution-based Exp.2	NAD, ATP
VVITAPSSTAPMFVMGVNEEK	P00358	Glyceraldehyde-3-phosphate dehydrogenase 2	Solution-based Exp.1A	NAD, ATP
YDSTHGRYAGEVSHDDK	P00358	Glyceraldehyde-3-phosphate dehydrogenase 2	Gel-based Exp. 2	NAD, ATP
YDSTHGRYAGEVSHDDK	P00358	Glyceraldehyde-3-phosphate dehydrogenase 2	Gel-based Exp. 2	NAD, ATP
YDSTHGRYAGEVSHDDK	P00358	Glyceraldehyde-3-phosphate dehydrogenase 2	Gel-based Exp. 2	NAD, ATP
VINDAFGIEEGLM(ox)TTVHSM(ox)TATQK	P00358	Glyceraldehyde-3-phosphate dehydrogenase 2;Glyceraldehy	Solution-based Exp.2	NAD, ATP
VINDAFGIEEGLMTTVHSMTATQK	P00358	Glyceraldehyde-3-phosphate dehydrogenase 2;Glyceraldehy	Solution-based Exp.1B	NAD, ATP
KVVITAPSSTAPM(ox)FVM(ox)GVNEEK	P00359	Glyceraldehyde-3-phosphate dehydrogenase 2;Glyceraldehy	Gel-based Exp. 1	NAD, ATP
KVVITAPSSTAPM(ox)FVM(ox)GVNEEK	P00359	Glyceraldehyde-3-phosphate dehydrogenase 2;Glyceraldehy	Solution-based Exp.1B	NAD, ATP
KVVITAPSSTAPMFVMGVNEEK	P00359	Glyceraldehyde-3-phosphate dehydrogenase 2;Glyceraldehy	Solution-based Exp.1B	NAD, ATP
VVITAPSSTAPM(ox)FVM(ox)GVNEEK	P00359	Glyceraldehyde-3-phosphate dehydrogenase 2;Glyceraldehy	Gel-based Exp. 1	NAD, ATP
VVITAPSSTAPM(ox)FVM(ox)GVNEEK	P00359	Glyceraldehyde-3-phosphate dehydrogenase 2;Glyceraldehy	Solution-based Exp.1B	NAD, ATP
VVITAPSSTAPM(ox)FVM(ox)GVNEEK	P00359	Glyceraldehyde-3-phosphate dehydrogenase 2;Glyceraldehy	Solution-based Exp.2	NAD, ATP
VVITAPSSTAPM(ox)FVMGVNEEK	P00359	Glyceraldehyde-3-phosphate dehydrogenase 2;Glyceraldehy	Solution-based Exp.1B	NAD, ATP

VVITAPSSTAPMFVMGVNEEK	P00359	Glyceraldehyde-3-phosphate dehydrogenase 2;Glyceraldehy	Solution-based Exp.1B	NAD, ATP
LTGM(ox)AFRVPTVDVSVVDLTVK	P00359	Glyceraldehyde-3-phosphate dehydrogenase 2;Glyceraldehy	Solution-based Exp.2	NAD, ATP
LTGMAFRVPTVDVSVVDLTVK	P00359	Glyceraldehyde-3-phosphate dehydrogenase 2;Glyceraldehy	Solution-based Exp.1B	NAD, ATP
GVLGYTEDAVVSSDFLGDSHSSIFDAS AGIQLSPK	P00359	Glyceraldehyde-3-phosphate dehydrogenase 3	Gel-based Exp. 2	NAD, ATP
IATYQERDPANLPWGSSNVDIAIDSTG VFK	P00359	Glyceraldehyde-3-phosphate dehydrogenase 3	Gel-based Exp. 2	NAD, ATP
IATYQERDPANLPWGSSNVDIAIDSTG VFK	P00359	Glyceraldehyde-3-phosphate dehydrogenase 3	Gel-based Exp. 2	NAD, ATP
VINDAFGIEEGLM(ox)TTVHSLTATQK	P00359	Glyceraldehyde-3-phosphate dehydrogenase 3	Gel-based Exp. 2	NAD, ATP
VINDAFGIEEGLM(ox)TTVHSLTATQK	P00359	Glyceraldehyde-3-phosphate dehydrogenase 3	Solution-based Exp.1B	NAD, ATP
VINDAFGIEEGLM(ox)TTVHSLTATQK	P00359	Glyceraldehyde-3-phosphate dehydrogenase 3	Solution-based Exp.2	NAD, ATP
VINDAFGIEEGLMTTVHSLTATQK	P00359	Glyceraldehyde-3-phosphate dehydrogenase 3	Solution-based Exp.1A	NAD, ATP
VINDAFGIEEGLMTTVHSLTATQK	P00359	Glyceraldehyde-3-phosphate dehydrogenase 3	Solution-based Exp.1B	NAD, ATP
IDGYSGPELGELM(ox)EK	P38088	Glycine--tRNA ligase 1, mitochondrial	Solution-based Exp.2	ATP
VESHLLNM(ox)SQDDLASK	P38088	Glycine--tRNA ligase 1, mitochondrial	Solution-based Exp.2	ATP
HQVDLSGSGTNNM(ox)GSNGAPK	P51996	GTP-binding protein YPT32/YPT11	Solution-based Exp.2	GTP
AM(ox)YTLSAQDEVFSLAFSPNRYWLA AATATGIK	P38011	Guanine nucleotide-binding protein subunit beta-like pr	Solution-based Exp.1B	GTP
DGEIM(ox)LWNLAAK	P38011	Guanine nucleotide-binding protein subunit beta-like pr	Solution-based Exp.2	GTP
SDVM(ox)SVDIDKK	P38011	Guanine nucleotide-binding protein subunit beta-like pr	Solution-based Exp.1B	GTP

YAIDM(ox)TEQARQGK	P31539	Heat shock protein 104	Solution-based Exp.2	ATP
NVAAGCNPM(ox)DLRRGSQVAVEK	P19882	Heat shock protein 60, mitochondrial	Solution-based Exp.2	ATP
HVFSATQLAAM(ox)FIDK	P32589	Heat shock protein homolog SSE1	Solution-based Exp.2	ATP
DAGTIAGLNVLRIINEPTAAAIAYGLDK	P10591	Heat shock protein SSA1	Gel-based Exp. 2	ATP
EPNRSINPDEAVAYGAAVQAAILTGDESSK	P10591	Heat shock protein SSA1	Gel-based Exp. 2	ATP
LIPRNSTIPTK	P10591	Heat shock protein SSA1	Gel-based Exp. 2	ATP
LVTDYFNGK	P10591	Heat shock protein SSA1	Gel-based Exp. 2	ATP
NQLESIAYSLK	P10591	Heat shock protein SSA1	Gel-based Exp. 2	ATP
VNDAVVTVPAYFNDSQRQATK	P10591	Heat shock protein SSA1	Gel-based Exp. 2	ATP
NQAAM(ox)NPANTVFDAK	P10592	Heat shock protein SSA2	Solution-based Exp.2	ATP
RLIGRNFNDPEVQGDM(ox)K	P10592	Heat shock protein SSA2	Solution-based Exp.2	ATP
DLLLLDVAPLSLGVGM(ox)QGDM(ox)FGIVVPRNTTVPTIK	P11484	Heat shock protein SSB1	Solution-based Exp.2	ATP
ENTLLGEFDLK	P11484	Heat shock protein SSB1	Gel-based Exp. 2	ATP
SQIDEVVLVGGSTRIPK	P11484	Heat shock protein SSB1	Gel-based Exp. 2	ATP
SSNITISNAVGRLSSEEIEK	P11484	Heat shock protein SSB1	Gel-based Exp. 2	ATP
M(ox)VNQAEEFK	P40150	Heat shock protein SSB1;Heat shock protein SSB2	Solution-based Exp.2	ATP
DAGLSTSDISEVLLVGGMSRMPK	P0CS90	Heat shock protein SSC1, mitochondrial	Solution-based Exp.1B	ATP
ESEPM(ox)EVDEDDSK	P15705	Heat shock protein STI1	Solution-based Exp.2	ATPase inhibitor, HSP90, HSP 70 binding
LM(ox)SFPEAIADCNK	P15705	Heat shock protein STI1	Solution-based Exp.2	ATPase inhibitor, HSP90, HSP 70 binding
ELM(ox)QQIENFEK	P04807	Hexokinase-2	Solution-based Exp.2	ATP

SAEDASEFVGVGSIAAGGRYDNLVNM(ox)FSEASGK	P07263	Histidine--tRNA ligase, mitochondrial	Solution-based Exp.2	ATP
GENVDPEIATYNEM(ox)VEDLFEQLAK	P06101	Hsp90 co-chaperone Cdc37	Solution-based Exp.2	Hsp90, kinases
ALGIM(ox)ALLDEGETDWK	P00817	Inorganic pyrophosphatase	Solution-based Exp.1B	Mg, Metal, Diphosphate
ALGIM(ox)ALLDEGETDWK	P00817	Inorganic pyrophosphatase	Solution-based Exp.2	Mg, Metal, Diphosphate
ALGIMALLDEGETDWK	P00817	Inorganic pyrophosphatase	Solution-based Exp.1B	Mg, Metal, Diphosphate
ENNIFNM(ox)VVEIPRWTNAK	P00817	Inorganic pyrophosphatase	Solution-based Exp.1B	Mg, Metal, Diphosphate
M(ox)SNIDFQYDDSVK	P09436	Isoleucine--tRNA ligase, cytoplasmic	Solution-based Exp.2	ATP
ELDTIRNM(ox)EIWK	P06168	Ketol-acid reductoisomerase, mitochondrial	Solution-based Exp.2	NAD, Mg, metal ion
STGNFM(ox)TLEQTVEK	P26637	Leucine--tRNA ligase, cytoplasmic	Solution-based Exp.2	ATP
AAAEGVANLHLDEATGEM(ox)VSK	P15180	Lysine--tRNA ligase, cytoplasmic	Solution-based Exp.2	ATP
LVGELEDTCINPTFIFGHPQM(ox)M(ox)SPLAK	P15180	Lysine--tRNA ligase, cytoplasmic	Solution-based Exp.2	ATP
RYLDLIM(ox)NK	P15180	Lysine--tRNA ligase, cytoplasmic	Solution-based Exp.2	ATP
DNQTSADAVEDGQNTGAM(ox)SHAFIK	Q08601	Metacaspase-1	Solution-based Exp.2	N/A
VSVGDLRYM(ox)LTGLGEK	P53141	Myosin light chain 1	Solution-based Exp.2	Ca, metal ion
M(ox)SDSPADGLDLALK	Q07551	NADPH-dependent alpha-keto amide reductase	Solution-based Exp.2	NADP
NTWGSQLVDQIM(ox)DQVVTK	P53184	Nicotinamidase	Solution-based Exp.2	nicotinamide, metal, Zinc
NTWGSQLVDQIMDQVVTK	P53184	Nicotinamidase	Solution-based Exp.2	nicotinamide, metal, Zinc
VEEAVLQSQGK	P40010	Nuclear GTP-binding protein NUG1	Gel-based Exp. 2	GTP
SSM(ox)QIDNAPTPHNTPASVLNPSYLK	P25293	Nucleosome assembly protein	Solution-based Exp.2	DNA

GFGYAGSPFHRVIPDFM(ox)LQGGDFT AGNGTGGK	P14832	Peptidyl-prolyl cis-trans isomerase	Solution-based Exp.2	CsA
MPQTVEWSK	P38013	Peroxiredoxin type-2	Solution-based Exp.2	N/A
GLINDPDM(ox)NNSFQINK	P14843	Phospho-2-dehydro-3-deoxyheptonate aldolase, phenylalan	Solution-based Exp.2	PEP
AQLPAGSNGLM(ox)IDYSHGNSNK	P32449	Phospho-2-dehydro-3-deoxyheptonate aldolase, tyrosine-i	Solution-based Exp.2	PEP
ALENPTRPFLAILGGAK	P00560	Phosphoglycerate kinase	Gel-based Exp. 2	ATP
ALLDEVVK	P00560	Phosphoglycerate kinase	Gel-based Exp. 2	ATP
ALLDEVVK	P00560	Phosphoglycerate kinase	Gel-based Exp. 2	ATP
EGIPAGWQGLDNGPESRK	P00560	Phosphoglycerate kinase	Gel-based Exp. 2	ATP
ELQSLLGK	P00560	Phosphoglycerate kinase	Gel-based Exp. 2	ATP
GVEVVLPVDFIIADAFSADANTK	P00560	Phosphoglycerate kinase	Gel-based Exp. 2	ATP
IQLIDNLLDK	P00560	Phosphoglycerate kinase	Gel-based Exp. 2	ATP
ISHVSTGGGASLELLEGK	P00560	Phosphoglycerate kinase	Gel-based Exp. 2	ATP
ITSNQRIVAALPTIK	P00560	Phosphoglycerate kinase	Gel-based Exp. 2	ATP
LFAATVAK	P00560	Phosphoglycerate kinase	Gel-based Exp. 2	ATP
LSVQDLDLK	P00560	Phosphoglycerate kinase	Gel-based Exp. 2	ATP
SSAAGNTVIIGGGDTATVAK	P00560	Phosphoglycerate kinase	Gel-based Exp. 2	ATP
VLENTEIGDSIFDK	P00560	Phosphoglycerate kinase	Gel-based Exp. 2	ATP
YSLAPVAK	P00560	Phosphoglycerate kinase	Gel-based Exp. 2	ATP
FNTYRRSFDVPPPPIDASSPFSQK	P00950	Phosphoglycerate mutase 1	Gel-based Exp. 2	NA, ATP
GQQEAARAGELLK	P00950	Phosphoglycerate mutase 1	Gel-based Exp. 2	NA, ATP
GQQEAARAGELLK	P00950	Phosphoglycerate mutase 1	Gel-based Exp. 2	NA, ATP
HLEGISDADIAK	P00950	Phosphoglycerate mutase 1	Gel-based Exp. 2	NA, ATP
HLEGISDADIAK	P00950	Phosphoglycerate mutase 1	Gel-based Exp. 2	NA, ATP
LSRAIQTANIALEK	P00950	Phosphoglycerate mutase 1	Gel-based Exp. 2	NA, ATP

LSRAIQTANIALEK	P00950	Phosphoglycerate mutase 1	Gel-based Exp. 2	NA, ATP
LSRAIQTANIALEK	P00950	Phosphoglycerate mutase 1	Gel-based Exp. 2	NA, ATP
LVLVRHGQSEWNEK	P00950	Phosphoglycerate mutase 1	Gel-based Exp. 2	NA, ATP
NLFTGWVDVK	P00950	Phosphoglycerate mutase 1	Gel-based Exp. 2	NA, ATP
TVM(ox)IAAHGN(de)SLRGLVK	P00950	Phosphoglycerate mutase 1	Gel-based Exp. 1	NA, ATP
TVM(ox)IAAHGNSLRGLVK	P00950	Phosphoglycerate mutase 1	Gel-based Exp. 1	NA, ATP
TVM(ox)IAAHGNSLRGLVK	P00950	Phosphoglycerate mutase 1	Gel-based Exp. 2	NA, ATP
TVM(ox)IAAHGNSLRGLVK	P00950	Phosphoglycerate mutase 1	Solution-based Exp.1A	NA, ATP
TVM(ox)IAAHGNSLRGLVK	P00950	Phosphoglycerate mutase 1	Solution-based Exp.1B	NA, ATP
TVM(ox)IAAHGNSLRGLVK	P00950	Phosphoglycerate mutase 1	Solution-based Exp.2	NA, ATP
TVMIAAHGNSLRGLVK	P00950	Phosphoglycerate mutase 1	Gel-based Exp. 1	NA, ATP
TVMIAAHGNSLRGLVK	P00950	Phosphoglycerate mutase 1	Solution-based Exp.1A	NA, ATP
TVMIAAHGNSLRGLVK	P00950	Phosphoglycerate mutase 1	Solution-based Exp.1B	NA, ATP
VYPDVLYTSK	P00950	Phosphoglycerate mutase 1	Gel-based Exp. 2	NA, ATP
VYPDVLYTSK	P00950	Phosphoglycerate mutase 1	Gel-based Exp. 2	NA, ATP
VYPDVLYTSK	P00950	Phosphoglycerate mutase 1	Gel-based Exp. 2	NA, ATP
YVDPNVLPETESLALVIDRLLPYWQDVIAK	P00950	Phosphoglycerate mutase 1	Gel-based Exp. 2	NA, ATP
FGPIVSASLEK	P04147	Polyadenylate-binding protein, cytoplasmic and nuclear	Gel-based Exp. 2	RNA
GQNCLDIDGIMDNPRK	P36048	Pre-mRNA-splicing factor SNU114	Solution-based Exp.2	GTP
VSLLTPDIGAVYSGM(ox)GPDYRVLVDK	P23639	Proteasome component Y7	Solution-based Exp.2	ATP dependent proteolytic Activity
HLQILQM(ox)PEK	P16550	Protein APA1;5,5-P-1,P-4-tetraphosphate phosphorylase;A	Solution-based Exp.2	ATP
EQFPLFAIHDM(ox)TEDLK	P17967	Protein disulfide-isomerase	Solution-based Exp.2	N/A

SENRTVSQGEM(ox)TSVHEPSTEEM(ox)AK	P46949	Protein FYV8	Solution-based Exp.2	N/A
GLNNTSVGEEK	Q04409	Putative glucokinase-2	Gel-based Exp. 1	ATP
AIGVQNAYFPM(ox)FVSSRVLEK	P38708	Putative proline--tRNA ligase YHR020W	Solution-based Exp.2	ATP
M(ox)FNLSVENPLGSDHPK	P38708	Putative proline--tRNA ligase YHR020W	Solution-based Exp.2	ATP
DYETHRVATTGEWDK	P06169	Pyruvate decarboxylase isozyme 1	Gel-based Exp. 2	TPP, Mg, Metal ion
DYETHRVATTGEWDK	P06169	Pyruvate decarboxylase isozyme 1	Gel-based Exp. 2	TPP, Mg, Metal ion
GSIDEQHPRYGGVYVGTLSKPEVK	P06169	Pyruvate decarboxylase isozyme 1	Gel-based Exp. 2	TPP, Mg, Metal ion
GSIDEQHPRYGGVYVGTLSKPEVK	P06169	Pyruvate decarboxylase isozyme 1	Gel-based Exp. 2	TPP, Mg, Metal ion
GSIDEQHPRYGGVYVGTLSKPEVK	P06169	Pyruvate decarboxylase isozyme 1	Gel-based Exp. 2	TPP, Mg, Metal ion
GYKPVAVPARTPANAAVPASTPLK	P06169	Pyruvate decarboxylase isozyme 1	Gel-based Exp. 2	TPP, Mg, Metal ion
GYKPVAVPARTPANAAVPASTPLK	P06169	Pyruvate decarboxylase isozyme 1	Gel-based Exp. 2	TPP, Mg, Metal ion
IRNATFPGVQM(ox)K	P06169	Pyruvate decarboxylase isozyme 1	Gel-based Exp. 1	TPP, Mg, Metal ion
IRNATFPGVQM(ox)K	P06169	Pyruvate decarboxylase isozyme 1	Gel-based Exp. 2	TPP, Mg, Metal ion
IRNATFPGVQM(ox)K	P06169	Pyruvate decarboxylase isozyme 1	Gel-based Exp. 2	TPP, Mg, Metal ion
KLIDLTQFPAFVTPM(ox)GK	P06169	Pyruvate decarboxylase isozyme 1	Gel-based Exp. 2	TPP, Mg, Metal ion
LIDLTQFPAFVTPM(ox)GK	P06169	Pyruvate decarboxylase isozyme 1	Gel-based Exp. 2	TPP, Mg, Metal ion
LLQTPIDM(ox)SLKPNDAESEK	P06169	Pyruvate decarboxylase isozyme 1	Gel-based Exp. 2	TPP, Mg, Metal ion

LLQTPIDM(ox)SLKPNDAESEK	P06169	Pyruvate decarboxylase isozyme 1	Gel-based Exp. 2	TPP, Mg, Metal ion
LLQTPIDM(ox)SLKPNDAESEK	P06169	Pyruvate decarboxylase isozyme 1	Gel-based Exp. 2	TPP, Mg, Metal ion
LLTTIADAAK	P06169	Pyruvate decarboxylase isozyme 1	Gel-based Exp. 2	TPP, Mg, Metal ion
LLTTIADAAK	P06169	Pyruvate decarboxylase isozyme 1	Gel-based Exp. 2	TPP, Mg, Metal ion
NIVEFHSDHM(ox)K	P06169	Pyruvate decarboxylase isozyme 1	Gel-based Exp. 2	TPP, Mg, Metal ion
YGM(ox)GTAASRSSAM(ox)TEYFK	P16387	Pyruvate dehydrogenase E1 component subunit alpha, mito	Solution-based Exp.2	Pyruvate, TPP, ATP*
AGLNIVRM(ox)NFSHGSYEYHK	P00549	Pyruvate kinase 1	Gel-based Exp. 2	ATP
GVNLPGTDVDLPALSEK	P00549	Pyruvate kinase 1	Gel-based Exp. 2	ATP
GVNLPGTDVDLPALSEK	P00549	Pyruvate kinase 1	Gel-based Exp. 2	ATP
GVNLPGTDVDLPALSEK	P00549	Pyruvate kinase 1	Gel-based Exp. 2	ATP
IENQQGVNNFDEILK	P00549	Pyruvate kinase 1	Gel-based Exp. 2	ATP
IM(ox)YVDYK	P00549	Pyruvate kinase 1	Gel-based Exp. 2	ATP
KGDTYVSIQGFK	P00549	Pyruvate kinase 1	Gel-based Exp. 2	ATP
VTDGVM(ox)VARGDLGIEIPAPEVLAVQK	P00549	Pyruvate kinase 1	Gel-based Exp. 2	ATP
VTDGVM(ox)VARGDLGIEIPAPEVLAVQK	P00549	Pyruvate kinase 1	Gel-based Exp. 2	ATP
VTDGVM(ox)VARGDLGIEIPAPEVLAVQK	P00549	Pyruvate kinase 1	Solution-based Exp.2	ATP
TM(ox)EEDEEVLYK	P41920	Ran-specific GTPase-activating protein 1	Solution-based Exp.2	GTP
YYNAVNPFEFM(ox)EDVATAGK	P49723	Ribonucleoside-diphosphate reductase small chain 2	Solution-based Exp.2	cytidine 5'-diphosphate
DLEDIGAGDQGIMFGYATDETPEGLPLTILLAHK	P10659	S-adenosylmethionine synthase 1	Solution-based Exp.1B	ATP

NFDLRPGVLVK	P10659	S-adenosylmethionine synthetase 1	Gel-based Exp. 2	ATP
M(ox)LDINQFIEDK	P07284	Serine--tRNA ligase, cytoplasmic	Solution-based Exp.2	ATP
IEQFVITEPEK	P07284	Seryl-tRNA synthetase, cytoplasmic	Gel-based Exp. 2	ATP
TVPEQSRDADVDASQGASAGGLPDLG SLLGGGLGGLM(ox)NNPQLM(ox)QAA QK	Q12118	Small glutamine-rich tetratricopeptide repeat-containin	Solution-based Exp.2	Hsp70, regulates ATPase activity
IGLINVRVPLPFNVAK	P39692	Sulfite reductase [NADPH] flavoprotein component	Gel-based Exp. 2	NADP, FAD, FMN, Flavoprotein
IEEIGGLIKKVNTSVK	P32854	Syntaxin PEP12	Gel-based Exp. 2	N/A
VEM(ox)IGDEYFSFLDNCK	P39077	T-complex protein 1 subunit gamma	Solution-based Exp.2	ATP
GIDPM(ox)SLDVFAK	P39079	T-complex protein 1 subunit zeta	Solution-based Exp.2	ATP
HSSTM(ox)NNFHIRSK	Q06163	Telomerase reverse transcriptase	Solution-based Exp.2	DNA, Mg, Metal ion
LQETNPEEVPK	P35691	Translationally-controlled tumor protein homolog	Gel-based Exp. 2	N/A
M(ox)QSDTQEANDIVTLANLQYNGST PADAFETK	Q12273	Transposon Ty1-PR2 Gag-Pol polyprotein;Capsid protein;T	Solution-based Exp.2	ATP, DNA, Mg, Metal, RNA, Zinc
M(ox)QTSNVLILGLK	P22515	Ubiquitin-activating enzyme E1 1	Solution-based Exp.2	ATP
APQAAGVIHNDLM(ox)NTFILAQVM(ox)K	P38219	Uncharacterized GTP-binding protein OLA1	Solution-based Exp.2	ATP
FFPSM(ox)ISPDWEPSIIPSNK	P53929	Uncharacterized protein YNL108C	Solution-based Exp.2	N/A
THSTYAFESNTNSVAASQMRNALNK	P32861	UTP--glucose-1-phosphate uridylyltransferase	Solution-based Exp.1B	UTP, Mg, Metal ion
ATGQM(ox)TERETM(ox)ETEVIK	P21576	Vacuolar protein sorting-associated protein 1	Solution-based Exp.2	GTP
GSQAM(ox)VM(ox)VEEK	P21576	Vacuolar protein sorting-associated protein 1	Solution-based Exp.2	GTP
EAVDM(ox)EFGTGAVK	P07806	Valine--tRNA ligase, mitochondrial	Solution-based Exp.2	ATP
GIYPPINVLPSLSRLM(ox)K	P16140	V-type proton ATPase subunit B	Solution-based Exp.2	ATP

ALERDVDLIESM(ox)K	P22203	V-type proton ATPase subunit E	Solution-based Exp.2	ATPase activity
TADLYAAM(ox)GLSK	P32527	Zuotin	Solution-based Exp.2	DNA, ATPase activity

References

[1] C. M. Dobson, A. Šali and M. Karplus, *Angewandte Chemie International Edition* 1998, *37*, 868-893.

[2] J. A. Schellman, *Annual Review of Biophysics and Biophysical Chemistry* 1987, *16*, 115-137.

[3] C. L. Araya, D. M. Fowler, W. Chen, I. Muniez, J. W. Kelly and S. Fields, *Proceedings of the National Academy of Sciences* 2012, *109*, 16858-16863.

[4] C. M. Dobson, *Nature* 2003, *426*, 884-890.

[5] C. M. Dobson, *Trends in Biochemical Sciences* 1999, *24*, 329-332.

[6] a) J. A. Schellman, *Biopolymers* 1975, *14*, 999-1018; b) J. A. Schellman, *Biopolymers* 1976, *15*, 999-1000.

[7] J. M. Sanchez-Ruiz, *Biophysical Chemistry* 2007, *126*, 43-49.

[8] T. T. Waldron and K. P. Murphy, *Biochemistry* 2003, *42*, 5058-5064.

[9] A. Marco in *Strategies for Boosting the Accumulation of Correctly Folded Recombinant Proteins Expressed in Escherichia coli*, *Vol. 752* Eds.: A. F. Hill, K. J. Barnham, S. P. Bottomley and R. Cappai), Humana Press, 2011, pp. 1-15.

[10] C. A. Minetti, P. L. Privalov and D. P. Remeta, *Proteins in Solution and at Interfaces: Methods and Applications in Biotechnology and Materials Science* 2013, 139-177.

[11] P. L. Privalov and N. N. Khechinashvili, *Journal of Molecular Biology* 1974, *86*, 665-684.

[12] E. Chautard, N. Thierry-Mieg and S. Ricard-Blum, *Pathologie Biologie* 2009, *57*, 324-333.

[13] S. Fields and O. K. Song, *Nature* 1989, *340*, 245-246.

[14] H. Zhu, M. Bilgin, R. Bangham, D. Hall, A. Casamayor, P. Bertone, N. Lan, R. Jansen, S. Bidlingmaier, T. Houfek, T. Mitchell, P. Miller, R. A. Dean, M. Gerstein and M. Snyder, *Science* 2001, *293*, 2101-2105.

[15] A.-C. Gavin and C. Hopf, *Drug Discovery Today: Technologies* 2006, *3*, 325-330.

[16] K. H. Young, *Biology of Reproduction* 1998, *58*, 302-311.

[17] A. McFedries, A. Schwaid and A. Saghatelian, *Chemistry & biology* 2013, *20*, 667-673.

[18] S.-E. Ong, M. Schenone, A. A. Margolin, X. Li, K. Do, M. K. Doud, D. R. Mani, L. Kuai, X. Wang, J. L. Wood, N. J. Tolliday, A. N. Koehler, L. A. Marcaurelle, T. R. Golub, R. J. Gould, S. L. Schreiber and S. A. Carr, *Proceedings of the National Academy of Sciences* 2009, *106*, 4617-4622.

[19] S.-E. Ong, B. Blagoev, I. Kratchmarova, D. B. Kristensen, H. Steen, A. Pandey and M. Mann, *Molecular & Cellular Proteomics* 2002, *1*, 376-386.

[20] B. Lomenick, R. Hao, N. Jonai, R. M. Chin, M. Aghajan, S. Warburton, J. Wang, R. P. Wu, F. Gomez, J. A. Loo, J. A. Wohlschlegel, T. M. Vondriska, J. Pelletier, H. R. Herschman, J. Clardy, C. F. Clarke and J. Huang, *Proceedings of the National Academy of Sciences* 2009, *106*, 21984-21989.

[21] a) P.-F. Liu, D. Kihara and C. Park, *Journal of Molecular Biology* 2011, *408*, 147-162; b) Y. Chang, J. P. Schlebach, R. A. VerHeul and C. Park, *Protein Science* 2012, *21*, 1280-1287; c) C. Park and S. Marqusee, *Nat Meth* 2005, *2*, 207-212.

[22] R. J. T. Corbett, F. Ahmad and R. S. Roche, *Biochemistry and Cell Biology* 1986, *64*, 953-961.

[23] a) M. J. Evans, A. Saghatelian, E. J. Sorensen and B. F. Cravatt, *Nat Biotech* 2005, *23*, 1303-1307; b) Y. Manabe, M. Mukai, S. Ito, N. Kato and M. Ueda, *Chemical Communications* 2010, *46*, 469-471.

[24] a) D. Harder and D. Fotiadis, *Nat. Protocols* 2012, *7*, 1569-1578; b) S. A. Sundberg, *Current Opinion in Biotechnology* 2000, *11*, 47-53.

[25] M. Vedadi, F. H. Niesen, A. Allali-Hassani, O. Y. Fedorov, P. J. Finerty, G. A. Wasney, R. Yeung, C. Arrowsmith, L. J. Ball, H. Berglund, R. Hui, B. D. Marsden, P. Nordlund, M. Sundstrom, J. Weigelt and A. M. Edwards, *Proceedings of the National Academy of Sciences* 2006, *103*, 15835-15840.

[26] R. Aebersold and M. Mann, *Nature* 2003, *422*, 198-207.

[27] a) T. Nilsson, M. Mann, R. Aebersold, J. R. Yates, A. Bairoch and J. J. M. Bergeron, *Nature Methods* 2010, *7*, 681-685; b) B. Domon and R. Aebersold, *Nat Biotech* 2010, *28*, 710-721; c) L. M. F. de Godoy, J. V. Olsen, J. Cox, M. L. Nielsen, N. C. Hubner, F. Fröhlich, T. C. Walther and M. Mann, *Nature* 2008, *455*, 1251-1254; d) S.-E. Ong and M. Mann, *Nature Chemical Biology* 2005, *1*, 252-262.

[28] S. Ghaemmaghami, M. C. Fitzgerald and T. G. Oas, *Proceedings of the National Academy of Sciences of the United States of America* 2000, *97*, 8296-8301.

[29] M. Fitzgerald and G. West, *Journal of the American Society for Mass Spectrometry* 2009, *20*, 1193-1206.

[30] C. N. Pace in *[14]Determination and analysis of urea and guanidine hydrochloride denaturation curves*, *Vol. Volume 131* (Ed. S. N. T. C. H. W. Hirs), Academic Press, 1986, pp. 266-280.

[31] K. D. Powell and M. C. Fitzgerald, *Biochemistry* 2003, *42*, 4962-4970.

[32] B. A. Howard, Z. Zheng, M. J. Campa, M. Z. Wang, A. Sharma, E. Haura, J. E. Herndon, M. C. Fitzgerald, G. Bepler and E. F. Patz, *Lung Cancer* 2004, *46*, 313-323.

[33] P. D. Dearmond, G. M. West, V. Anbalagan, M. J. Campa, E. F. Patz and M. C. Fitzgerald, *Journal of Biomolecular Screening* 2010, *15*, 1051-1062.

[34] G. M. West, L. Tang and M. C. Fitzgerald, *Analytical Chemistry* 2008, *80*, 4175-4185.

[35] a) G. M. West, C. L. Tucker, T. Xu, S. K. Park, X. Han, J. R. Yates and M. C. Fitzgerald, *Proceedings of the National Academy of Sciences* 2010, *107*, 9078-9082; b) P. D. DeArmond, Y. Xu, E. C. Strickland, K. G. Daniels and M. C. Fitzgerald, *Journal of Proteome Research* 2011, *10*, 4948-4958; c) E. C. Strickland, M. A. Geer, D. T. Tran, J. Adhikari, G. M. West, P. D. DeArmond, Y. Xu and M. C. Fitzgerald, *Nat. Protocols* 2013, *8*, 148-161.

[36] M. Miyagi and T. Nakazawa, *Analytical Chemistry* 2008, *80*, 6481-6487.

[37] R. L. Nagel and Q. H. Gibson, *Journal of Biological Chemistry* 1971, *246*, 69-&.

[38] G. M. West, C. L. Tucker, T. Xu, S. K. Park, X. Han, J. R. Yates, III and M. C. Fitzgerald, *Proceedings of the National Academy of Sciences of the United States of America* 2010, *107*, 9078-9082.

[39] K. Kawahara, A. G. Kirshner and C. Tanford, *Biochemistry* 1965, *4*, 1203-&.

[40] C. N. Pace, D. V. Laurents and J. A. Thomson, *Biochemistry* 1990, *29*, 2564-2572.

[41] C. N. Pace, *Critical Reviews in Biochemistry* 1975, *3*, 1-43.

[42] S. H. C. Ip and G. K. Ackers, *Journal of Biological Chemistry* 1977, *252*, 82-87.

[43] R. W. Henkens, B. B. Kitchell, S. C. Lottich, P. J. Stein and T. J. Williams, *Biochemistry* 1982, *21*, 5918-5923.

[44] D. Andersson, P. Hammarstrom and U. Carlsson, *Biochemistry* 2001, *40*, 2653-2661.

[45] A. N. Szilagyi and M. Vas, *Folding & Design* 1998, *3*, 565-575.

[46] J. S. Valentine, P. A. Doucette and S. Z. Potter in *Copper-zinc superoxide dismutase and amyotrophic lateral sclerosis*, *Vol. 74* 2005, pp. 563-593.

[47] G. Mei, N. Rosato, N. Silva, R. Rusch, E. Gratton, I. Savini and A. Finazziagro, *Biochemistry* 1992, *31*, 7224-7230.

[48] Y. Xu, I. N. Falk, M. A. Hallen and M. C. Fitzgerald, *Analytical Chemistry* 2011, *83*, 3555-3562.

[49] J. A. Tainer, V. A. Roberts and E. D. Getzoff, *Current Opinion in Biotechnology* 1992, *3*, 378-387.

[50] A. D. Weston and L. Hood, *Journal of Proteome Research* 2004, *3*, 179-196.

[51] M. J. Campa, M. Z. Wang, B. Howard, M. C. Fitzgerald and E. F. Patz, *Cancer Research* 2003, *63*, 1652-1656.

[52] J. Lee, *Archives of Pharmacal Research* 2010, *33*, 181-187.

[53] F. R. Howard B.A., Campa M.J., Rabbani Z.N., Vujaskovic Z., Wang X.-F., Patz Jr. E.F., *Cancer Research* 2005, *65*, 8853-8860.

[54] K. Okuzawa, B. Franzén, J. Lindholm, S. Linder, T. Hirano, T. Bergman, Y. Ebihara, H. Kato and G. Auer, *ELECTROPHORESIS* 1994, *15*, 382-390.

[55] G. Chen, T. G. Gharib, D. G. Thomas, C.-C. Huang, D. E. Misek, R. D. Kuick, T. J. Giordano, M. D. Iannettoni, M. B. Orringer, S. M. Hanash and D. G. Beer, *Proteomics* 2003, *3*, 496-504.

[56] C. S. Gan, P. K. Chong, T. K. Pham and P. C. Wright, *Journal of Proteome Research* 2007, *6*, 821-827.

[57] W. W. Wu, G. Wang, S. J. Baek and R.-F. Shen, *Journal of Proteome Research* 2006, *5*, 651-658.

[58] K. Gevaert, J. Van Damme, M. Goethals, G. R. Thomas, B. Hoorelbeke, H. Demol, L. Martens, M. Puype, A. Staes and J. Vandekerckhove, *Molecular & Cellular Proteomics* 2002, *1*, 896-903.

[59] L. M. Wolfe, U. Veeraraghavan, S. Idicula-Thomas, S. Schürer, K. Wennerberg, R. Reynolds, G. S. Besra and K. M. Dobos, *Molecular & Cellular Proteomics* 2013, *12*, 1644-1660.

[60] Y. Nozaki, *Methods in enzymology* 1972, *26 PtC*, 43-50.

[61] L. J. Licklider, C. C. Thoreen, J. Peng and S. P. Gygi, *Analytical Chemistry* 2002, *74*, 3076-3083.

[62] J. Cox and M. Mann, *Nat Biotech* 2008, *26*, 1367-1372.

[63] P. DeArmond, G. West, H.-T. Huang and M. Fitzgerald, *Journal of the American Society for Mass Spectrometry* 2011, *22*, 418-430.

[64] P.-F. Liu and C. Park, *Journal of Molecular Biology* 2012, *422*, 403-413.

[65] N. Seidler in *Dynamic Oligomeric Properties*, *Vol. 985* Springer Netherlands, 2013, pp. 207-247.

[66] *Prudent Practices in the Laboratory: Handling and Disposal of Chemicals*, The National Academies Press, 1995, p.

[67] A. Shevchenko, H. Tomas, J. Havlis, J. V. Olsen and M. Mann, *Nat. Protocols* 2007, *1*, 2856-2860.

[68] M. Saraste, P. R. Sibbald and A. Wittinghofer, *Trends in Biochemical Sciences* 1990, *15*, 430-434.

[69] B. K. Szpikowska, J. M. Beechem, M. A. Sherman and M. T. Mas, *Biochemistry* 1994, *33*, 2217-2225.

[70] A. Fattoum, C. Roustan, D. Karoui, J. Feinberg, L.-A. Pradel, J. Gregoire and H. Rochat, *International Journal of Peptide and Protein Research* 1981, *17*, 393-400.

[71] A. Matte and L. T. J. Delbaere in *ATP-binding Motifs*, *Vol.* John Wiley & Sons, Ltd, 2001.

[72] B. M. Gorovits and P. M. Horowitz, *Journal of Biological Chemistry* 1995, *270*, 28551-28556.

[73] J. K. Myers, C. N. Pace and J. M. Scholtz, *Protein Science* 1995, *4*, 2138-2148.

[74] C. R. Bagshaw, *Journal of Cell Science* 2001, *114*, 459-460.

[75] A. Seybert, D. J. Scott, S. Scaife, M. R. Singleton and D. B. Wigley, *Nucleic acids research* 2002, *30*, 4329-4338.

[76] M. Ünlü, M. E. Morgan and J. S. Minden, *ELECTROPHORESIS* 1997, *18*, 2071-2077.

Biography

Duc T. Tran was born on June 15, 1986 in Thai Binh, a town in Vietnam. She attended the Vietnam National University of Science in 2004 and received a bachelor degree in Biotechnology in 2008. Tran joined the PhD program in Biochemistry at Duke University - Graduate School of Arts and Science in 2008. She has published a work entitled *"Slow Histidine H/D exchange protocol for Thermodynamic Analysis of Protein folding and stability using Mass spectrometry"* on The Journal of Analytical Chemistry in 2012. Tran coauthors an article on Nature Protocols Journal entitled *"Thermodynamic Analysis of Protein-ligand binding interactions in complex biological mixtures using the Stability of Protein from Rates of Oxidation"* in 2013. Tran received the Vietnam Education Foundation Fellowship from 2008-2013. She is a member of the American Society for Mass Spectrometry since 2009.